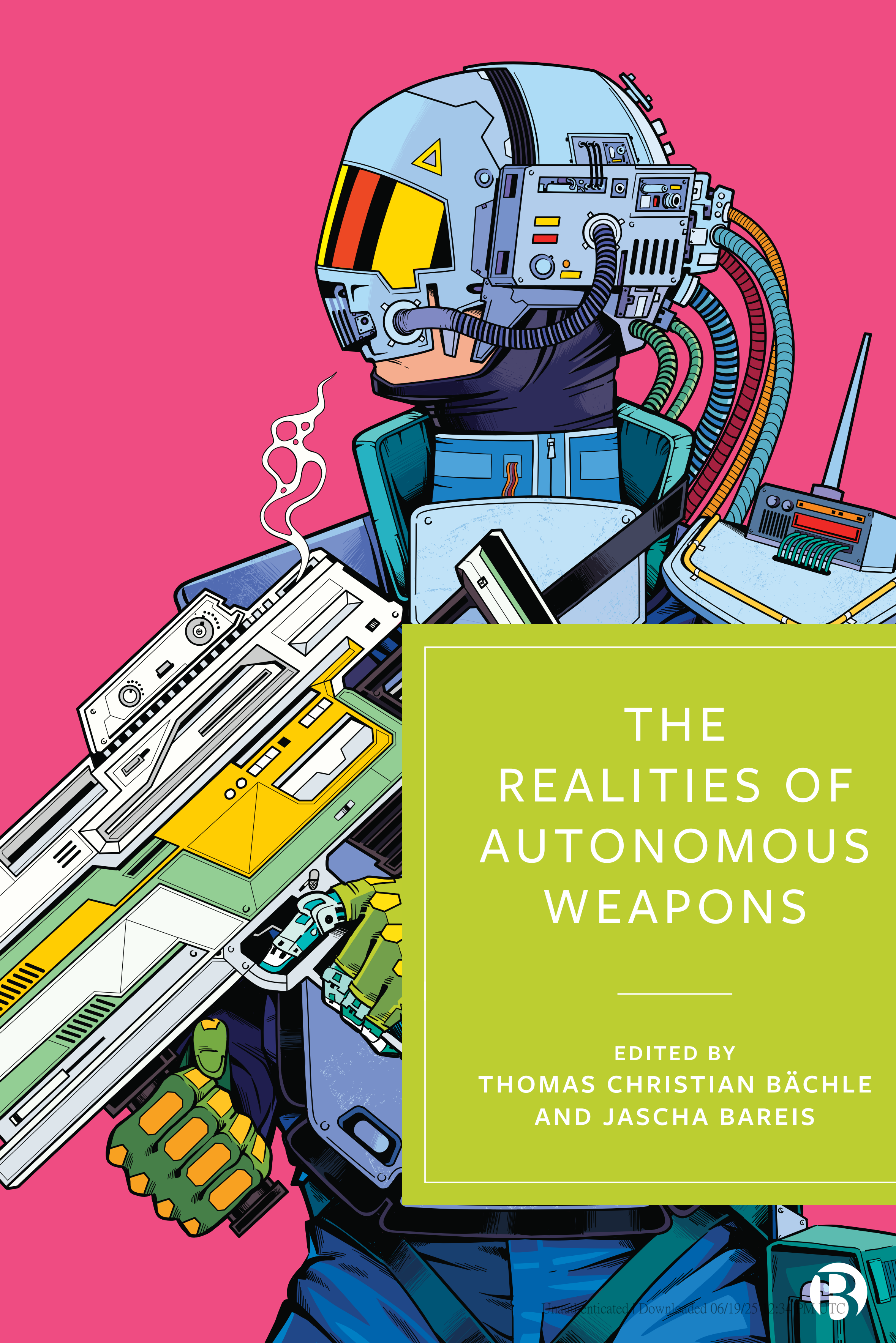


THE REALITIES OF AUTONOMOUS WEAPONS

EDITED BY
THOMAS CHRISTIAN BÄCHLE
AND JASCHA BAREIS

THE REALITIES OF AUTONOMOUS WEAPONS

Edited by
Thomas Christian Bächle and
Jascha Bareis

BRISTOL
UNIVERSITY
PRESS

First published in Great Britain in 2025 by

Bristol University Press
University of Bristol
1-9 Old Park Hill
Bristol
BS2 8BB
UK
t: +44 (0)117 374 6645
e: bup-info@bristol.ac.uk

Details of international sales and distribution partners are available at bristoluniversitypress.co.uk

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN 978-1-5292-5109-8 paperback
ISBN 978-1-5292-3720-7 ePub
ISBN 978-1-5292-3719-1 ePdf

Cover design: Andy Ward
Front cover image: iStock/Man_Half-tube
Bristol University Press uses environmentally responsible print partners.
Printed and bound in Great Britain by CPI Group (UK) Ltd, Croydon, CR0 4YY

FSC
www.fsc.org
MIX
Paper | Supporting responsible forestry
FSC® C013604

Contents

List of Figures and Tables

Figures

Tables

From the Reel to the Real

From Network-Centric Warfare to Autonomous Warfighting Networks

Notes on Contributors

Thomas Christian Bächle is Head of the Digital Society Research Programme at the Alexander von Humboldt Institute for Internet and Society (HIIG) in Berlin and researcher at the University of Bonn. His areas of research include human/machine interaction, technological materialities, interfaces and agency; mobile media, surveillance, robotics, affective computing and simulation technologies.

Jascha Bareis is a political scientist and researcher at the Institute for Technology Assessment and Systems Analysis (ITAS) at the Karlsruhe Institute of Technology (KIT) and associate researcher at the Humboldt Institute for Internet and Society (HIIG) in Berlin. He investigates political communication and power at the crossroads of technology governance, hype and BigTech. His work has been published in journals and blogs like *Science, Technology & Human Values*, *Big Data & Society* and *Verfassungsblog*.

Peter Behrbohm is a Berlin-based artist and architect exploring future conflicts and utopian societies through collaborative, site-specific and context-based research. He designs operational obstacles and fictitious institutions. Many of his works are surgically intervening public spaces, routines and discourses, often aiming at revealing the constructedness of realities and demonstrating the possibilities of other forms of coexistence.

Ingvild Bode is Associate Professor at the Center for War Studies, University of Southern Denmark. Her research focuses on processes of normative and policy change, especially with regard to the use of force. Her work has been published in the *European Journal of International Relations*, *Review of International Studies*, *International Studies Review* and other journals. She is the Principal Investigator of the European Research Council (ERC)-funded project AutoNorms (2020–2025), investigating how autonomous weapon systems change norms.

Christoph Ernst is a research assistant (tenured) at the Department of Media Studies of the University of Bonn. His main research interests are diagrammatic thinking and interface studies (especially cognition and tacit knowledge), future media and theory of sociotechnical imagination, aesthetics and theory of visual media (especially film and television), and media and the military (especially autonomous weapon systems). Among his publications are *Diagramme zwischen Metapher und Explikation – Studien zur Medien- und Filmästhetik der Diagrammatik* (transcript, 2021); *Media Futures: An Introduction* (Palgrave Macmillan, 2021); *Tech | Demo* (edited with Julia Eckel and Jens Schröter, Universi, 2023). Further information can be found here: www.christoph-ernst.com.

Jens Hälterlein is a researcher at the Department of Media Studies at Paderborn University and scientific coordinator of the 'Meaningful Human Control – Autonomous Weapon Systems between Regulation and Reflection' (MEHUCO) project funded by the German Federal Ministry of Education and Research. He has been publishing on the social dimension of security technologies since 2011 – among others, within the framework of the EU project VideoSense. From 2020 to 2022, he was Principle Investigator of the 'Artificial Intelligence and Civil Security' project funded by the Fritz Thyssen Foundation. He has many years of research experience in the fields of science and technology studies, surveillance studies, critical security studies and governmentality studies.

Teresa Heffernan is Professor of English Language and Literature at Saint Mary's University, Halifax, NS. Her current research is on the science and fiction of robotics and AI. Her books include the edited collection *Cyborg Futures: Cross-disciplinary Perspectives on Artificial Intelligence and Robotics* (Palgrave Macmillan, 2019); *Veiled Figures: Women, Modernity, and the Spectres of Orientalism* (University of Toronto Press, 2016); and *Post-Apocalyptic Culture: Modernism, Postmodernism, and the Twentieth-Century Novel* (University of Toronto Press, 2008). She runs the following website: https://socialrobotfutures.com

Rebecca L. Jones has published on women's representation in comics, video games and fantasy adaptations. She is interested in exploring the representation of minority characters within genre fiction, film, television, comics and games. For her master's degree, she studied the adaptation of Shakespeare within comics, Marvel films and science-fantasy novels, and in her doctorate she examined the representational evolution of female-presenting artificial creations in American and British live-action films.

Xiran Liu is a master's student in Communication Science at the University of Zurich. She holds a bachelor's degree in journalism from Northwestern University and a master's degree in Social Policy from the London School of Economics and Political Science (LSE). Her current research interest is the public communication about emerging technologies such as generative AI on social media.

Shimona Mohan is Associate Researcher on Gender & Disarmament and Security & Technology at the United Nations Institute for Disarmament Research (UNIDIR) in Geneva, Switzerland. Her previous work includes stints at the Observer Research Foundation (ORF), the Think20 Secretariat under India's G20 presidency in 2023, the Carnegie Endowment for International Peace, and the Ministry of External Affairs as well as NITI Aayog under the Government of India. Her areas of research include the multifarious intersections of security, technology (especially AI and cybersecurity), gender and disarmament.

Jeremy Packer is Professor at the Institute for Communication, Culture, Information, and Technology at the University of Toronto. He is the author or co-author of *The Prison House of the Circuit: Politics of Control from Analog to Digital* (University of Minnesota Press, 2023); *Killer Apps: War, Media, Machine* (Duke University Press, 2020); and *Mobility Without Mayhem: Cars, Safety, and Citizenship* (Duke University Press, 2008). He is also the editor or co-editor of *Foucault, Cultural Studies, and Governmentality* (State University of New York Press, 2003); *Thinking with James Carey: Essays on Communications, Transportation, History* (Peter Lang Publishing, 2006); *Secret Agents: Popular Icons Beyond James Bond* (Peter Lang Publishing, 2009); and *Communication Matters: Materialist Approaches to Media, Mobility, and Networks* (Routledge, 2012).

Joshua Reeves is Associate Professor at the School of Communication at Oregon State University. He is the author or co-author of *The Prison House of the Circuit: Politics of Control from Analog to Digital* (University of Minnesota Press, 2023); *Killer Apps: War, Media, Machine* (Duke University Press, 2020); and *Citizen Spies: The Long Rise of America's Surveillance Society* (New York University Press, 2017). He serves as an associate editor of *Surveillance & Society* and is a director of the Surveillance Studies Network.

Jennifer Rooke served 24 years as an intelligence officer in the US Air Force upon graduation from the Air Force Academy until her military retirement in 2013 at the rank of colonel. She worked across a breadth of organizational levels, in operational squadrons and headquarters staff. These included assignments as a director of operations and a commander of intelligence squadrons, as well as staff positions within Headquarters Air Force, the National Security Agency, US European Command in Germany and the Office of the Secretary of Defense. She also held senior intelligence officer positions at US

Air Forces in Europe, headquartered in Germany, and at NATO Allied Joint Force Command Brunssum, headquartered in the Netherlands.

Elke Schwarz is Reader (Associate Professor) in Political Theory at Queen Mary University London. She works at the intersection of ethics of war and ethics of technology with an emphasis on unmanned and autonomous/intelligent military technologies and their impact on warfare.

Bernhard Seidl is Senior Lecturer in Japanese Studies at the Department of East Asian Studies, University of Vienna. His research interest is focused on exploring metapragmatic discourse and current social discourses in Japan through the use of text corpora. An avid fan of *rakugo*, the Japanese art of humorous storytelling, he has also dabbled in research on modern performing arts in Japan.

Lucy Suchman is Professor Emerita of the Anthropology of Science and Technology at Lancaster University. Before taking up that post, she was a principal scientist at Xerox's Palo Alto Research Center (PARC), where she spent 20 years as a researcher. Her current research extends her longstanding critical engagement with the fields of artificial intelligence and human-computer interaction to the domain of contemporary militarism. She is concerned with the question of whose bodies are incorporated into military systems, how and with what consequences for social justice, and the possibility for a less violent world.

Jinyu Wang studies Communication Design in Saarbrücken, Germany. Her work focuses on providing visual insights into the challenges of humanity in modern society, aiming to provoke greater attention and deeper thought on complex moral issues.

Jutta Weber is a Science and Technology Studies scholar and Professor for Media, Culture and Society at Paderborn University. Her research focuses on computational technoscience culture(s), asking how and for whom the (non)human actors work. She has been visiting professor at Uppsala (Sweden), Vienna (Austria) and Twente (the Netherlands). She is lead of the competence network 'Meaningful Human Control: Autonomous Weapon systems between Regulation and Reflection'; you can find more information here: www.juttaweber.eu; www.mehuco.de

Johannes Weilandt works as a visual artist in Berlin. He studied fine art at the Weißensee Kunsthochschule Berlin and graduated as a master's student in 2022. His mostly serial drawings have been nominated for several awards and are represented in various exhibitions.

Acknowledgements

The editors wish to thank Christoph Ernst, Elisa Birkmann and Dana Abuosba for their help and support. The open access funding of the volume and the allowances for the artists in this volume has generously been enabled by the Institute for Technology Assessment and Systems Analysis (ITAS) at the Karlsruhe Institute for Technology (KIT). Special thanks go to Linda Nierling and Reinhard Heil of the ITAS research group 'Digital Technologies and Societal Change' (DigIT) for their support of the project, and to Beate Meiswinkel from ITAS for her efforts to enable all the financial transactions. The editors also wish to thank the Humboldt Institute for Internet and Society in Berlin, the Department of Media Studies at the University of Bonn and the 'Meaningful Human Control' research network for their support.

The Realities of Autonomous Weapons: Hedging a Hybrid Space of Fact and Fiction

Jascha Bareis and Thomas Christian Bächle

Introduction

The development of autonomous weapon systems (AWS) – at times also bearing the 'lethal' label under the acronym LAWS – has been subject to intense discussions for years. Numerous political, academic or legal institutions and actors debate the consequences and risks that arise from these technologies, particularly their ethical, social and political implications, with many calling for strict regulation or even a global ban. Despite this public prominence and the perceived consequentiality of these weapons, it often remains surprisingly unclear which technologies are evoked by the term AWS and what they are capable of. AWS can refer to drones, flight carriers, unmanned aerial, ground or maritime vehicles, robots and robot soldiers or cyberweapons such as computer viruses.

This uncertainty comes despite (or maybe because of) the fact that there are numerous definitions that try to specify the term either functionally ('once activated', autonomous weapons 'can select and engage targets without further intervention by an operator': US Department of Defense, 2023: 21) or conceptually (derived from the theorization of autonomous systems, artificial intelligence [AI] or machine learning [ML]). Definitions leave plenty of room for different types of technologies and – combined with the much wider discussions on AI – potentials and projections on future developments. Besides this terminological ambiguity, it also remains inherently vague in what sense and to what degree these systems can be characterized as autonomous at all. Even though the development of automated capabilities is undoubtedly advancing (Scharre, 2018; Schwarz, 2018; Packer and Reeves, 2020), with an ever-decreasing degree of human agency and ways to intervene, fully autonomous weapons that are completely beyond human control and for this reason feared by many largely represent a conceptual possibility rather than an actual military reality.

These ambiguities result in wide gaps of meaning, which are in turn filled with imaginations – a common practice for new technologies and AI in particular (Suchman, 2023). Potential realities can fulfil an important role, as they are tools to transfer expert knowledge into other fields of society, including journalism, policy making, research, education and democratic decision-making processes. Hence, the ideas on the functionality of AWS and their consequences are inspired and shaped by imaginaries on military, national and technological futures. They include geopolitical scenarios, ethical questions, national policies or science fiction. In security and military policies, these interconnections between different realities are even utilized as a methodology – for example, as 'red teaming' – which means applying creative fictional accounts of potential futures to inform actual decision making (The Red Team, 2021). Another application is war gaming, a method of foreseeing future military scenarios originating at least as far back as the 19th century, but adapted to contemporary technological and media environments, including virtual reality and AI-based simulations using large language models (Goecks and Waytowich, 2024).

The premise of autonomous weapons, seen as entertaining a hybrid space of their own, invites the exploration of their concomitant myriad realities. The rationale of the book maintains that the realities in question can only be understood by acknowledging the constant and complex dynamic between the actual technological developments and the visions and virtual scenarios that are associated with them. It is exactly in this context of uncertainty – in which imagination, possibility and fiction are conflated – that autonomous weapons become highly consequential. They provoke emotions, discourses, agitations, (re)actions, investments, competition, policies or technological and military blueprints.

Publications on the topic of autonomous weapons often focus on their legal, political or ethical ramifications (for example, Bhuta et al, 2016; Krishnan, 2016), a first-order level of assessing these technologies, with some works also discussing their unique representations (Graae and Maurer, 2021), and the way we witness and experience them (Bousquet, 2018; Richardson, 2024). The foundation of these works is also based on the different realities outlined earlier. Introducing another way of analyzing the realities of autonomous weapons, this volume puts forward a second-order level approach: an ethical problem, for example, is not framed only as such, that is, along the lines of posing the following normative question: 'Which moral questions arise with automated killing machines?' The ethical problem, in the approach suggested here, is instead to understand it as a contributing factor that helps to construct, disseminate and maintain a specific understanding of lethal AWS in popular culture, politics, journalism or research. In short, ethical discourses co-create the realities of their object. For this reason, the perspectives taken in this book foreground the different realities of AWS and, in turn, aim at informing the existing debates about their (often implicit) underlying assumptions.

This introductory chapter of the book first sketches out the technological and political developments towards an ever-increasing automation of military machinery. These developments are theorized as both constitutive as well as performative to encompass the dynamics and different understandings of AWS around the globe in theory and practise. Subsequently, the chapter offers six reflections on these realities that help hedge and consolidate the dynamic meanings of AWS, which tend to receive so much attention in public, military and regulatory arenas. The chapter concludes with an overview of the book's structure and a brief summary of the individual contributions.

Approaching the realities of autonomous weapons

The idea of automatic or self-directed weapon systems can be traced back (at least) to the cybernetic paradigm of the 1940s (Galison, 1994). However, in military history, the final phase of the Cold War in the late 1980s and the First Gulf War in 1991 can be seen as the key moment towards today's discourses on AWS, since the Cold War also saw the first philosophical examination of 'intelligent' war machines (de Landa, 1991). Against the background of various ideas on 'post-industrial' warfare (for example, Echevarria and Shaw, 1992; Toffler and Toffler, 1993), the digitalization of information and communication infrastructures of the US Armed Forces has been characterized as a 'Revolution in Military Affairs' (Cohen 1996) and considered as a phase of disruptive technological developments. Around the same time, the paradigm of network-centric warfare emerged, which defined the standards for a new form of warfare based on the idea to achieve permanent information dominance through rigorous networking of all forms of military systems, including both human and technical agents (see Ernst's chapter in this volume; Cebrowski, 2005; Bousquet, 2022).

Another milestone in the political and military ambitions to intensify the development of AWS – especially in the field of robotics (Lin et al, 2008) – is marked by the terrorist attacks conducted on 11 September 2001 in the US and their aftermath (Singer, 2010). Most notably, weaponized drones such as the US MQ-9 Reaper (by General Atomics) or the X-47 series (by Northrop Grumman) were rapidly developed during a time that was labelled 'the War on Terror'. Subsequently, the notion of an 'Age of Autonomous Systems' in warfare (Worcester, 2015) or calls to urgently start 'preparing for war in the robotic age' (Work and Brimley, 2014) have emerged in recent years. These visions were strongly driven by the military utilization of more recent forms of AI such as machine learning tools or artificial neural networks (Cummings, 2017). The latest iteration of an innovative AI-related hype (Bareis et al, 2023) – at the time of writing – has been featured via the concept of 'generative AI', which has also entered both the vocabularies and the imaginations of military industries (Goecks and Waytowich, 2024) and armed forces (Lushenko and Carter, 2024).

The realities of AWS also include the dynamic between fact and fiction. They are often influenced by popular culture and inspired by more general assumptions about AI and its relationships to the human in the broadest sense, echoing tropes such as the substitution of humans by machines, or the risks of intelligent machines that are no longer subjected to human control. These realities are thus shaped by a mix of intentional framing and larger sociocultural narratives that act on a discursive rather than an individual level. A well-known position is the idea that autonomous weapons can be seen as more fair or just (Arkin, 2009). The obvious ethical and critical questions are 'What enables the framing of an instrument for surveillance and killing as an inherently ethical instrument? What kind of sociopolitical rationale underpins such a framing?' (cf. Schwarz, 2018). In other words, the framing of ethicality is produced by but also produces particular realities of autonomous weapons, for example as contested moral arbiters or dystopian slaughter machines.

The book also touches upon conceptual approaches to autonomous warfare technologies, shaping the ways in which they are modelled, developed or advertised in their interactions with humans. Well-known examples for this in the context of regulating autonomous weapons are the often normatively utilized descriptors of 'meaningful human control' on the part of humans and 'autonomous' on the part of machines. It is necessary to stress that both bear meanings that are constructed and constructive rather than descriptive (Bächle, 2023). These dynamic meanings prove to be particularly challenging in legal assessments that require a normative stance. Scholars have started to challenge the apparent consensus that human judgement is to be treated as a legal requirement in the context of AWS, questioning the commonly shared foundations of regulation rooted in ideas such as explainability, accountability, dignity or the principle of humanity. When AI-enabled technologies are compared to other types of weapons, one issue is still not settled: 'If we want better human control, we need to explain why' (Lubell, 2023). Interestingly, this condition is not verbalized as strongly for other types of weapons systems (such as anti-personnel landmines), which can be equally harmful but are not met with a similar concern, involving explicit human oversight. This is not to say that weapons of mass destruction (biological, chemical, radiological or nuclear) are any less consequential. However, their development and actual employment largely pre-dates international regulatory frameworks (most notably that under the United Nations) and presents a different historical context. Also, the world might simply have had more time to get used to them. A technology such as AWS, seen by many as genuinely novel, arguably triggers a heightened sense of uncertainty and attention. Paired with the complexities of a multicentred geopolitical context and competing media realities, the differing perception of urgency and threat – this is one of the book's assumptions – is attributed to the fluctuating nature of the realities of autonomous weapons.

Conceptually, the realities of autonomous weapons are connected – but not identical – to what Jasanoff and Kim (2015) call sociotechnical imaginaries. According to their definition, sociotechnical imaginaries are 'collectively held, institutionally stabilised, and publicly performed visions of desirable futures … and supportive of advances in science and technology' (Jasanoff and Kim, 2009: 120). Sociotechnical imaginaries inform realities of autonomous weapons, especially in the field of state discourse and political communication, as communication in the public arena presupposes a shared understanding among larger social groups. In these public arenas, imaginaries point to, as Jasanoff (2015) argues, 'positive visions of social progress … [and], tacitly or explicitly, with the obverse – shared fears of harms that might be incurred through invention and innovation, or of course the failure to innovate' (Jasanoff, 2015: 4–5).

However, the understanding of realities in this book goes further. The very idea of AWS is closely interwoven with military histories and current hopes and developments towards machine intelligence and the possibilities of human agency. Historically, AWS' military imaginations, contexts and discourses are continuous and dynamic developments that cannot be tied to one singular event or technical breakthrough; rather, they can only be understood through the lens of their technical precursors and the shared norms and values of their time. The understandings that are associated with AWS also vary geographically, which means they cannot be reduced to one emblematic representation – often US- and Eurocentric – such as of killer robots or drone swarms (Arquilla and Ronfeldt, 2000; Coeckelbergh, 2011). The realities of autonomous weapons take into account popular aesthetics, fictions, policies and corporate discourses that can differ significantly cross-culturally.

This overlap between the technological paradigms and their larger societal and cultural manifestations show that AWS are not only shared and understood in clearly articulated visions or imaginaries. They are characterized by mediation, frictions and hybridity that create a reality of their own. For example, efforts to predict future military threats, conflict scenarios and simulations under the condition of *potential* technological advancements is equivalent to the creation of 'as if' realities. These *virtual* – potentially innate – realities of AWS shape the *actual* debates on their ethical and legal ramification. They inform ways of representation in public discourse and the basis of political decision making today. For this reason, this book argues that AWS are *created* as objects, while at the same time drawing 'distinctions between life and death, human and machine, culture and technology' (Karppi, Böhlen and Granata, 2018: 107).

Media technologies play an important role in this (for example, Hoskins and O'Loughlin, 2015), which is not limited to merely representing warfare and warfare technologies. Baudrillard (1995) famously commented that the Gulf War in 1991 was not taking place. He described its reality as not bound to the battleground and constituted by actual combat operations, but as coming into effect via mediated, mainly televised form, broadcasting live into the living rooms of North American and European citizens. Mediatized and mediated warfare creates simulations of war, representations that do not presuppose actual events. The Gulf War points to the virtuality of war. It was not necessary for it to take place to become a reality in the TV living rooms. The idea of mediated warfare became even more prevalent after the terrorist attacks on 11 September, 2001 in the US: the paradigmatic importance of drones – in particular, the claim of high-precision drone strikes – for the supposedly new forms of warfare is interrelated with normative questions associated with these weapons systems (Krasmann and Weber, 2015). From a technical standpoint, drones are not necessarily autonomous systems, but rather remote-controlled robots (unmanned combat aerial vehicles), which are able to independently perform specific subtasks such as surveillance and reconnaissance. Nevertheless, drones have made a reality *imaginable*, in which technical autonomous systems are able to perform kill decisions independently of human control (Maurer and Graae, 2021). Their prominent representation in the media also established a particular aesthetics of drone images (Richardson, 2020; see also the artwork by Weilandt in this volume). Drones are emblematic of a detached and distant view, reinforcing the narrative of technologically assisted clean, precise and efficient forms of warfare against the enemy – favourably depicted as 'terrorist vermin' in the 2000s (Sarasin, 2006). In a more abstract sense, drones have thus been established as both real technologies and symbols for the imagination of an expectable future, in which fully autonomous combat robots are no longer a purely fictitious possibility (Elish, 2018).

The mediated realities of AWS have to be accounted for, especially given new media environments, which incorporate virtual reality, augmentation and digital forms of decentralized communication – and, lately, the rise of synthetically produced media with text and pictures through generative AI. This not only leads to a de facto convergence of military and entertainment media (Lenoir and Caldwell, 2018), when, for example, interfaces used to control drones are inspired and optimized by computer games and vice versa. But media forms themselves *co-create* the realities of warfare, often in a fuzzy overlap of temporalities and media spheres. The recent violent conflicts in Ukraine and the Middle East have highlighted the ways in which social media publics are targeted in propaganda wars (Rudloff and Appel, 2023), while public authorities try to engineer opinions in a desired fashion. The new media environments also enable first-person accounts of their experiences – evoking labels such as soldiers, terrorists, civilians and innocents – even allowing them to livestream *their own* reality of on-the-ground combat (Rarm, 2023). The use of generative AI makes it increasingly difficult to ascertain whether these accounts are authentic or fake (Antinori, 2019).

Despite these vast fields of AI applications in hybrid warfare, and somewhat paradoxically, the public perception of autonomous weapons – promoted by state actors, the militaries or industries – is often reduced to machinistic understanding of weapons: unmanned vehicles, missiles or drones. These materialistic imaginations reduce the broad range of conducting attacks to an underestimated field of digital and AI-enabled warfare (Shaw, 2016; Merrin, 2018). However, cyberattacks quite holistically aim at the manipulation or destruction of computer software or devices, which disrupt not only militaries but potentially also all aspects of our digital lives. 'Autonomous' computer viruses or cyberattacks do not just hit our capabilities to communicate, but potentially all mediated aspects of social reality and also the everyday material objects – the Internet of Things – that surround us (Arquilla, 2021). The manipulation of publics through misinformation, targeted leaks or the disruption of traditional media and journalism of media also thrives (Seib, 2021). This new media environment entails a power shift to platforms and private companies that are seen by many as responsible for moderating and regulating the content that they make accessible.

Acknowledging the overlap and conflation of fact and fiction, the real and the virtual, the truthful and the fake, the desired and the detested is the main conceptual baseline for the analyses presented in this book. It is established (and good) practice for current research to strongly focus on normative issues of legal and ethical regulation of AWS in order to inform policy makers, politicians, the military industry and civil society. However, the realities of autonomous weapons take a different, constructivist route to this end. This volume interrogates different media, aesthetics, histories and visions, as well as geographical particularities with regard to *their own* realities. It aims to make explicit the tacit knowledge on AWS by calling into question their taken-for-granted preconditions and manifestations. In doing so, the book also seeks to make a contribution to the relevant normative debates with their legal and political implications.

To do justice to this conceptual angle, the volume features contributions from different academic disciplines, with each of them prioritizing a particular approach to and aspect of the wide range of different realities. To emphasize that cultural negotiations also play a major role in constructing the realities of AWS, each of the book's three sections is introduced by the work of an artist and their unique take on the phenomena in question.

The following six reflections pinpoint these complex realities of autonomous weapons by addressing common (mis)conceptions and by locating them within some of the larger contexts sketched above.

1. Autonomous weapons systems are perceived as clandestine technologies evoking curiosity and awe

AWS development is mostly classified. States conceal the latest technology advancements in the name of the national interest, with agencies and laboratories working on military innovations shielded from the public eye. Supremacy in weaponry power is trending high on many national and geostrategic security agendas (see, for example, Bächle and Bareis [2022] for a comparison of the US and China). It embodies a military and industrial striving for competitive advantage in a perceived arena of threat and rivalry. The urgency and legitimacy is derived from mobilizing a rhetoric of fierce international competition, thereby hailing technological innovation as a pillar of national resilience capabilities (Bareis and Katzenbach, 2022).

A prominent example is the US Defense Advanced Research Projects Agency (DARPA). DARPA was founded by President Eisenhower in 1958 and during its planning phase, it was initially coined the Special Projects Agency (Barber Associates, 1975: 59). It was created in reaction to the Soviet-induced Sputnik crisis among US political elites. Still today its aim is to 'formulate and coordinate "breakthrough" technologies and capabilities for national security' (DARPA, n.d.) together with academic research and industry. A self-assuring DARPA promotional video introduces the founding motif in 1958, which hails DARPA as being 'the initiator, not the victim of strategic technological surprises' (DARPAtv, 2018: 0:24).

Institutions like DARPA function as mission-oriented agencies (Mazzucato, 2011), which are legitimated by the imperative of state leadership, often at the cost of democratic processes. It is common that they trade transparency and public accountability for speed and secrecy in the name of the national interest. The role of public funding and the 'hidden Developmental State' (Block, 2008) with agencies such as DARPA (or its European equivalent, the Joint European Disruptive Initiative: JEDI Foundation, n.d.) have changed throughout the years to become more similar to a network of public-private partnerships. State agencies cooperate with major technology corporations contributing to military and intelligence imperatives. Some of these projects were famously leaked in the past, such as the common surveillance practices by the US, which were made public by former intelligence employee Edward Snowden (Lyon, 2015). Only after protests by Google employees did the public learn about plans to collaborate with the Pentagon under the name of Project Maven, which in 2018 incorporated the company's AI technology in order to analyse drone surveillance footage (Simonite, 2021; see also Heffernan's chapter in this volume). The idea of a hidden power structure gets also easily misused, for example by the first Trump administration and in its aftermath by utilizing a 'deep state' conspiracy theory (Horwitz, 2022).

The state's role in facilitating military innovation recedes and leaves a gap that is filled by the private sector, which often comes at the cost of ethical considerations. In the US, Silicon Valley is increasingly setting the agenda for military innovation and focuses especially on dual-use technology, driven by a bottom-up, neoliberal and corporate-led strategy: 'It flips these defense acquisition processes on their heads such that industry drives, rather than responds to, militaries' requirements for new capabilities' (Lushenko and Carter, 2024). Big technology companies and billionaires provide military infrastructure and start to set the constraints on the battlefields – for example, Elon Musk influenced conflict dynamics by deciding whether Ukraine could use the Starlink satellite network or not. Private stakeholders like Musk shape the country's military operations 'on the basis of *his* fears of crisis escalation' (Lushenko and Carter, 2024).

The concealing of state agencies and the rise of power of private companies in the name of the national interest leaves room for imaginations and rumour, exploiting a deep fascination with the inaccessible, clandestine – and seemingly powerful and out of control. The military industry thrives in this context of uncertainty. This fascination can be compared to the spectacle around the highly classified Manhattan Project (1942–1946), the US research programme for developing a nuclear bomb. Technology was hailed as a means to rule the world and even heralded a new epoch of the Anthropocene: the 'nuclear age' (Hughes, 2004).

2. Autonomous weapons trigger both fascination and horror – and subscribe to common historical narratives of technology and dominance

The development and portrayal of AWS strongly speaks to and exacerbates the existing hopes and fears around AI (Cave and Dihal, 2019). Building on the age-old fascination for the latest technological development, they are simultaneously emblematic of potentially devastating effects and scenarios playing with themes of dominance and chaos (see also Bode and Mohan investigating sentiments in the Indian public, or Jones analysing the stereotypes of female-presenting AWS in cinema history in this volume).

There are two historical narratives – one rotating around the concept of dominance and the other around enhancement and extension – that entertain sentiments of fascination and horror with technology. The first regards science and technology as ways to control and cultivate nature, essentially establishing both as distinct realms (Latour, 1993). Taming the natural environment and its unpredictable force (through droughts, floods or earthquakes) rationalizes technology as a necessary force to expand and maintain human civilization through domination (Nye, 2004). Industrialization and engineering projects such as the construction of dams or railway networks epitomize the 'technological conquest of matter' (Marx, 2000: 197). Overcoming the physical limits of nature and matter plays on the imagination of achieving the seemingly impossible (Beckert, 2016).

The second historical discourse more directly refers to contexts of military technology as forms of enhancement and extension in an array of different techniques. First and foremost, this refers to weapons technology which makes it possible to increase the distance between soldiers and also decrease the need to engage in direct body combat. It includes swords, cannons, bows and arrows, necessitating protection gear such as shields or body armour (cf. Diamond, 1997). Another technique is the effort to enhance the biological capabilities of soldiers, a notorious example of which is the use of the methamphetamine Pervitin in the Second World War (Rasmussen, 2011). The foundational ideas of optimizing military strategy (Von Clausewitz, 1942) are instantiated in cultural techniques such as war gaming, academic approaches to capture the dynamics of war empirically (Bousquet, Grove and Shah, 2020) or the computer-assisted simulation and prediction of military scenarios today (Cayirci et al, 2022).

In both historical narratives, technology entertains notions of power and (loss of) control, either taming nature or subjugating enemies by enhancing the soldier and its abilities. Technology represents social, even magical and sublime qualities (Appadurai, 1986), or can elicit horror or repulsion, running the risk of rendering the human obsolete – a destruction even beyond imagination (Anders, 2002). It is these histories in which the cultural portrayal of autonomous weapons is rooted and finds its expression. For example, science fiction films and public campaigns cater to doomsday scenarios that mobilize pictures of merciless and destructive machines. AWS are pictured as 'killer robots' (see the initiative Stop Killer Robots, n.d.) or 'slaughterbots' (see movie Autonomous Weapons, n.d.). The idea of AI going rogue, turning against its makers and humanity at large, is another common trope in the theme of loss of control and taming. Autonomous and human-like machines evoke fears of a lethal intelligence that could outsmart humans. The (real) opacity of these AI-based systems, which cannot be comprehended by the majority of people, fosters the idea of networked architectures making themselves independent and taking on a 'life' of their own. Certainly, a great deal of the intimidation evoked by the sublime aura of AWS is produced through the limitless force of human imagination, quickly crossing the boundaries of fact and fiction. Take motifs of a sinister HAL 9000 computer in *Space Odyssey*, the idea of a cybernetic android killer such as *The Terminator* or scenarios of killer drone swarms (also depicted in the video Slaughterbots – see above), which reverberates with Alfred Hitchcock's menacing motif in *The Birds*. These portrayals of fictional destructive lethal machinery are sustainably shaping the public and political perceptions of AWS and are contributing to a large extent to their popularity.

3. Autonomous weapons as tools in psychological warfare and strategic communication

Putting into perspective the current detrimental effects of AWS, it is certainly noteworthy that conventional firearms – at the time of writing in 2024 – inflict more harm and human suffering than AI-assisted military technologies. In the US alone, the latest complete data show that in 2021, 48,830 people died from gun-related violence (Gramlich, 2023), with some sources suggesting 40,871 deaths for 2024 (Gun Violence Archive, 2024). In Mexico, official numbers give 22,309 gun-related deaths in 2022 (Álvarez, 2023), and in South Africa, 8,388 deaths in 2021, with numbers on the rise, as between October and December 2022 alone, more than 7,500 people died through firearms (Kirsten, 2023 and Khumalo, 2023). In 2022 in the US alone, the firearm and ammunition industry was responsible for as much as $80.73 billion in total economic activity of the country (NSSF, 2022). In comparison, in the same year, the *global* military AI market size was substantially smaller, valued at $7.4 billion in 2022 (Grand View Research, n.d.). However, the market size of autonomous military weapons is rising, with an estimated compound annual growing rate of 10.4% for the coming years (Business Research Company, 2024). Pistols and rifles seem to be perceived as conventional, almost traditional, and are more accepted among the public, even being hailed in parts of pop culture as a prop of masculinity. They have been widely disseminated and decentralized in use around the globe for decades. Also, they are comparably low-tech engineered and remain largely unchecked in trade – despite a global Arms Trade Treaty, although one which has not been signed by nations with major production sites (Amnesty International, n.d.). To give a comparison, the elaboration and highly differentiated debates on the future of warfare, subsequent AWS risk scenarios, and the assessments of their ethical repercussions seem strangely detached from the scale of harm and violence caused by conventional guns. Politically and normatively, it is harder to draw attention to the risks and harms of contemporary conventional ('stupid') weapons, as they lack the nimbus of glitzy AI-enabled future warfare, even though their global use and trade is beyond any meaningful human control. This is a standard which is often raised in AWS regulatory debates.

The main difference seems to be that the rhetorical drumbeat in relation to AWS is already part of modern warfare and used as an effective tool in strategic communication. Suggestions of AI capabilities, woven into the political rhetoric of state actors, can be an influential vehicle in the deterrence of enemies (Johnson, 2020). The praise of AWS capabilities can also be understood as a means of psychological warfare, with the aim to clarify one's position in the geopolitical order and strategically contain, defend or strive for hegemonic aspirations (see Bächle and Liu's chapter in this volume). As argued elsewhere, a comparison between Chinese and US AWS imaginaries shows that '[military] AI is in both cases regarded as a means to realise these socio-political ideals, with supremacy achieved by technological prowess being a shared theme for both' (Bächle and Bareis, 2022: 7).

The political and symbolic communication of AWS is not only used for deterrence but also mirrors cultural particularities and changes of the material actuations on the battlefield. Besides the conceptual vagueness, the terminology applied in the discussions on AWS is commonly contextualized in larger narratives. Often, machine capabilities in weaponry allude to broadly known mythological and anthropomorphic references, or borrow motifs from popular culture, religious or historical tropes. For example, the US counter rocket, artillery and mortar (C-RAM) close-in weapon system Phalanx, which has been in service since the 1980s, takes its reference from the ancient Greek war practice, where spears units formed a phalanx formation in battle against the enemy. The C-RAM vulcan cannon can be mounted on ships, and, next to the Greek reference for its name, the Navy's crews gave the Phalanx systems the pet name R2-D2 because their appearance is reminiscent of the droid R2-D2 from the *Star Wars* films (Stoner, 2009). At the time of writing, recent examples of attempts to foreground a branding of AI in military contexts can be found in the employment of target recommender systems. The Israel Defense Forces (IDF) use AI in the military operations in Gaza following the terrorist attacks by Hamas on Israeli civilians on 7 October 2023. The employed AI system is called Habsora, the 'gospel' – which translates into 'holy message' in biblical terms, building on an AI-powered database called Lavender for target classification, scoring and subsequent bombing of alleged terrorists (Abraham, 2024). Also, Ukrainian forces use recommender systems in the war against Russian troops, the so-called Geographic Information System Art for Artillery (GIS ARTA, n.d.) for fire missions, also being coined by its Ukrainian developer Sherstyuk 'Uber for Artillery' (Bruno, 2022). GIS Arta speeds up artillery missions by sourcing real-time data 'from drones, targets reported by forward observers armed with cell phones, counter battery radars, and satellite-based imagery' (Zikusoka, 2023).

The references to different motifs and imaginaries are meant to reach objectives in political and public communication – but as a side effect complicate shared understandings of military AI and AWS in the wider public, academic or political arenas. Their meanings become loaded with associations and references that can lead to appeal, but also face the risk of further misguiding and mystifying the technical functionality of AWS. Technology is subjected to interpretations in a discursive realm that is already heavily loaded with emotions, normative positioning and geopolitical power striving.

As another side effect, the overemphasis on the imagined potentials of modern intelligent weapons shifts the focus away from the very conventional and often very 'stupid' weaponry such as rifles or pistols, as mentioned earlier, or mass-produced simplistic drones (for example, the Iranian-Russian cooperation to produce Shahed-136 drones to attack Ukraine; Bennett and Ilyushina, 2023). They pose a threat by way of sheer quantity and easy access, as they can also be manufactured or commissioned by nonstate actors. The use of AWS abilities in contemporary battlefields shows no monolithic or linear development to fully automated battle machines. Rather than taking over the entire range of tasks in identifying, tracking and eliminating enemy objects, the current automation wave resembles a mix in weapon systems and approaches. For example, in the Ukrainian war, civil improvised drones are mounted with hand grenades and bombs to be used by Ukrainian forces, alongside conventional rocket, artillery and missile (RAM) ammunition. Soviet equipment from the Cold War and improvised dual-use gadgets from the civil realm blend with the latest AI analytics and civilian spying (*The Economist*, 2024). This not only proves the need for or capability of improvisation, but also fulfils the purposes of political communication. The collaboration of the civil and the commercial sectors with the Ukrainian military forces can be promoted as a sign of union, symbolizing hope and resilience for a nation in an exceptional state of emergency.

4. Autonomous weapons epitomize the fluidity of violence

From the perspective of international relations, AWS can be seen as a continuation of a prerogative of violence that transcends national borders and acts as an event outside of temporal and spatial limitations. The 2001 US Bush doctrine of the 'War on Terror' declared the necessity of effectively intervening against terrorist groups, no matter where they are located. The rise of violent nonstate actors operating across borders brings with it the risk of unchecked dissemination of weaponry among warlords, terrorists and private armies. Military terrorist groups such as ISIS or al-Qaeda are part of the privatization of war as much as state mercenaries like US Blackwater fighting in the Second Gulf war or the Russian Wagner group operating on the African continent. Here, easy-access and high-quantity automatic weapons must be regarded as a particular threat in the hands of these actors, employing harmful technologies outside of regulatory frameworks. Building largely on dual-use components, and being to a large extent software-based, makes the dissemination easier (often in a downloadable, intangible form) and at the same time more difficult to trace compared to conventional weapons.

AWS make distinctions ever more obsolete as the interoperability of algorithms between different use domains easily conflates categories such as civilian and military, corporate and government, and private and public. Military decision-making tasks in the realm of selection and targeting are almost indistinguishable from recommender systems used on commercial entertainment or social media platforms digesting content. Similar conflations happen in the realm of 3D printing, which produce both commercial tools or weapon components, or in the field of exoskeletons that may either support workers in factories or enhance the physical capabilities of soldiers.

From the perspective of state sovereignty, the erosion and conflation of state-controlled violence calls for a technical-military apparatus that enforces spatial and temporal dominance as a reaction. Intelligence agencies of states like the US enforce a prerogative of worldwide surveillance, categorizing, tracking and eliminating potential enemies through air strikes. For example, Rooke argues in her chapter in this volume that the US Air Force and its declaration of 'air-mindedness' became a pivotal factor in a 'hierarchical ordering that places the US at the top of this dominant spatiotemporality'. From this perspective, AWS in the form of drones and or AI-enabled cyberattacks resemble a form of warfare that executes power through both writing and simplistic categorizing (enemy/ally, hostage/terrorist) – combined with the kinetic ability to execute lethal power anytime and anywhere. The prerogative of air-mindedness goes along with the power to make perpetrators, victims or injustices (dis)appear, as they happen far outside the auspices of international humanitarian law (IHL), human rights and public accountability. Rupka and Baggiarini argue that air warfare conducted through drones resembles a 'militarised gaze', which is 'both everywhere and nowhere, whilst its power successfully enables the rendering of populations into the terrain of state legibility and security so that they might become governable subjects' (2018: 13). With or without an official declaration of war, states can operate effectively in the geopolitical realm without taking accountability for their actions. Violence acts without having troops on the ground, causing difficulties for the normative international system to hold perpetrators accountable.

An example of the practice of automating enemyship is the conflict that started following the attack by Hamas on Israel on 7 October 2023. The use of the Lavender and Habsora systems by the IDF fuelled speculation on how target recommendation of alleged 'militant suspects' can be automated. According to statements that were reportedly made by IDF soldiers, recommender systems were used for enemy detection, scoring people in the Gaza Strip with a rating of between 1 and 100, expressing how likely it was that they were to be a militant (Abraham, 2024). Automating target classification was believed to play 'a critical role in building lists of individuals authorised to be assassinated' by airstrikes (Davies et al, 2023; Abu Elouf, 2025). The consequent massive bombing of the Gaza Strip in response to the Hamas attacks led to immense civilian suffering and human rights violations, raising issues of proportionality, enemy/civilian distinction and accountability of Israeli bombings in respect to IHL.

In more general terms, for the perception of AWS, this type of warfare practice underscores how the *automation* of target recognition alone involves the critical issues of surveillance, acceleration and dehumanization of war, and entails the risk of reducing human lives to a probability score cleared for killing. Paradoxically, the promise of increased precision and effectiveness of air strikes through surveillance and algorithmic scoring simultaneously provokes an increased psychological perception of threat and insecurity. Making the hitherto unknown tangible by transforming it via algorithmic analysis into a seemingly uncontrollable quantity of hostility has a potential impact on the likelihood of escalation. As Packer and Reeves state in their chapter in this volume, when reflecting on the recursive relationship between media technology, knowledge creation and the production of threat: 'When applied in a political or military context, this means that enemies will always be found; with positive feedback systems, there is no way to ultimately find and neutralize all enemies; the system's operation demands the constant discovery of new problems to solve.' Both psychologically and kinetically, AWS represent and accelerate the enactment of violence across time and space.

5. Autonomy in weapon systems emphasizes the necessity to thoroughly theorize artificial intelligence

The ongoing efforts to regulate autonomous weapons and the use of AI has not just underlined the need to properly define what makes an autonomous weapon system reall*y autonomous* or what is characteristic of an AI system that sets it apart from its technological precursors; in a more abstract sense, it also places a spotlight on the many, still remaining conceptual voids surrounding current debates on autonomous systems and AI.

The rise of AI, especially accelerated by a combination of machine learning (ML) data processing capabilities, more effective sensors and advanced infrastructure, has enabled weapon systems to operate with much less human intervention than the preceding technologies could. The allure of

AI has seemingly changed attributes from *automatic* into *autonomous* systems, which sparks epistemic but also regulatory confusion (Sauer, 2016). From a disciplinary standpoint, autonomy has always been a contested concept. Also, in technical and engineering discourses, it has become a widely used term, where it commonly evokes associations of independence, intelligence, self-governance, self-sufficiency, the ability to learn and adapt (for example, orientation in unknown, unstructured and dynamic environments) or the execution of self-determined decisions (Williams, 2015). Such functional viewpoints in engineering easily conflate understandings of autonomy, trust and responsibility from the viewpoint of human moral agency (see Schwarz's chapter in this volume). As a consequence, and problematically so, technical understandings are starting to be applied in the realm of human ethics, resulting in a mechanical weighing of human value similar to mathematical calculation and algorithmic optimization.

It seems common practice among military and political stakeholders to reinterpret the concept of autonomy and AI to particular means, which often comes at the cost of nullifying the conceptual or practical use of the term. A position paper submitted in 2018 to the Convention on Certain Conventional Weapons (CCW) negotiations in Geneva by the German delegation, for example, states the following: 'Having the ability to learn and develop self-awareness constitutes an indispensable attribute to be used to define individual functions or weapon systems as autonomous' (Permanent Representation of the Federal Republic of Germany to the Conference on Disarmament in Geneva, 2018). Tying 'self-awareness' to a definition of machine autonomy is absurd, for obvious reasons. However, it can have a rhetorical function at the negotiation table. In the same year, the Chinese delegation at the CCW defined a necessary feature of AWS with the following condition: 'once started there is no way to terminate the device' (CCW Group of Governmental Experts on LAWS, 2018: 1). This entertains the no less absurd scenario of an AI gone rogue, completely outside of human control. Partly due to the terminological confusion and strategic vagueness, the CCW negotiations have become gridlocked and are far from reaching a consensus that would honour IHL in a serious attempt to regulate the actual reality of AWS (see also Suchman's chapter in this volume). Overall, some public and military interpretations of autonomy in the AWS debate articulate sensationalist fiction and have succeeded in capturing not only public discourses (see Cave and Dihal, 2019; Campolo and Crawford, 2020), but also debates in research (Natale and Ballatore, 2020), and have found their way in the regulatory arena (see Bächle and Bareis, 2022).

In addition, more conceptually grounded notions of autonomy in automated warfare are no historically fixed constants, but are subject to change. Ernst, for example, argues in his chapter in this volume that rather than dealing with self-sufficient and autonomous battle machines such as drones, tanks or ships, autonomy in contemporary military visions is better understood as resilient *networks* between connected agents and infrastructures. Combat clouds engage in warfighting, highlighting the importance of communication hubs or real-time data analytics. Projects such as the European Future Combat Air System (FCAS) also point in this direction (see Hälterlein's chapter in this volume). These examples of recommender systems or combat clouds highlight the various elements in warfare that are increasingly automated and hence different from the idea of a self-sustained 'autonomous' battle machine, as already noted.

Contrasting with many of the prevalent approaches used in political science, law or philosophy, which understand autonomy as a distinct quality associated with the human condition, these examples also indicate that autonomy instead emerges performatively within social or material structures and is thus subject to cultural change and national differences (Haraway, 2006). It sheds light on the *mechanisms* that provoke what could be called 'autonomy effects'. This performative understanding of autonomy also helps to look past many of the thought experiments and fictions that consider a world in which machines will finally have acquired human-like abilities. It is not only important to unpack the metaphorical uses and the practices of how autonomy is made (Noorman and Johnson, 2014), but also makes visible the networked and automated infrastructures that underlie imaginations around AWS.

It is exactly this interpretative openness of the term 'autonomy' that predestines it to be applied in various contexts and with tailor-made meanings. The erosion of its semantic qualities not only calls for a thorough reflection of the premises used, but even more importantly for a general theorization of AI.

6. Autonomous weapons challenge our understanding of what is human and foreground the relationship between humans and machines

As part of the shift away from solely looking at the suggested autonomy of a distinct system, it is particularly necessary to assess the human/machine relationship. Conceptually, reality of the existing 'human-machine autonomies' (Suchman and Weber, 2016) – rather than autonomous machines – have important roots in cybernetic theory, establishing an analogy between humans and machines via a universally applicable analogy: 'The systems analogy, as well as the understanding of systems as goal directed and purposeful, is a central precondition for the idea of the 'autonomy' of so-called smart and intelligent (war) machines' (Suchman and Weber, 2016: 83–84).

While the human/machine systems analogy is a theoretical precondition of common ideas of autonomy and autonomous weapons – often drawing false equivalencies, as discussed in the previous section – it paradoxically also elicits the paradigmatic question on differentiating humans and machines. In the most basic terms, this means asking about the human element, whether it being part of 'the loop' or in 'meaningful control'. Imagining weapons necessarily entails imagining a version of the human. It concerns their role in the relation with machines, as in ethical, political or legal categories: when and how should a human be able to intervene, should a human necessarily be involved in the decision to kill another human and so on.

On a par with this, the military discourse on AWS is no longer purely technocentric, but moves towards both the human/machine relationship or even human-centricity. 'Manned/unmanned teaming', 'human augmentation' (UK Ministry of Defence/Bundeswehr Office for Defence Planning, 2021), or 'the enhanced soldier' (de Boisboissel and Le Masson, 2021) take into account and shape these technological, conceptual and strategic shifts. Augmentation has even been identified as the up-and-coming paradigm in discussions of autonomous weapons and military AI (cf. Favaro and Schwarz, 2022).

'The human' has always been present in a functional sense, because it is a vital – but often only pro forma – point of reference. Debates on political, legal or ethical debates on responsibility, dignity, intentionality and so on require a human to pin them on: as long as 'the human' as a function is formally in the picture, the otherwise autonomous machine seems more legitimate. However, it is high time to direct our attention to humans. This means that rather than solely discussing AWS as technical entities, we need to focus on human/machine interactions and relations, and the ways in which they extend human capabilities of taking action or decrease the levels of skill or competence. It also means acknowledging that fully autonomous systems are – even though they foster our fascination and horror – a rather skewed narrative.

The book's sections and individual contributions

The book's structure introduces three individual sections that engage with current realities of autonomous weapons. Each section analyses autonomous weapons from a particular trope of perspective: 1. Narratives and Theories, 2. Technologies and Materialities and 3. Politics and Ethics. The beginning of each section is introduced by an artist and their vision on autonomous weapons. The sectioning adheres to an analysis of the different meanings articulated across these domains that constitute the realities of AWS and powerfully influence how we perceive and engage with this technology.

Section I: Narratives and Theories

This section looks at cultural texts that are marked as fiction (for example, science-fiction films and novels) as well as those marked as nonfiction in research. Its goal is to analyse the potentials, risks, narratives and aesthetics that are associated with AWS:

- *ARTWORK*. «The Unreachable Myth: Killing Unknown Victims with Insensible Means by Unidentified Perpetrators for Unapparent Reasons». By Jinyu Wang, 2023
- Jennifer Rooke: 'The AI/lure of US Airpower: Imaginaries of Disruption in the Pursuit of Technological Superiority Since the Early 20th Century'
- Rebecca Jones: 'From *Maschinenmensch* to Robot Bubs: Female-Presenting Autonomous Weapons Systems in Live-Action Films from 1927–2022'
- Teresa Heffernan: 'Autonomous Weapons in Fiction and the Fiction of Autonomous Weapons'
- Ingvild Bode and Shimona Mohan: 'From the Reel to the Real: Narratives of Weaponized Artificial Intelligence Technologies in India'

The artwork «The Unreachable Myth» by Jinyu Wang opens the section by pointing to the functionality of AWS in an illustrative storytelling format, inspired by the sci-fi aesthetics from the 1970s. In 'The AI-Lure of US Airpower: Imaginaries of Disruption in the Pursuit of Technological Superiority since the Early 20th Century', Jennifer Rooke analyses the military imaginaries that shape the use of automated pattern and target recognition technologies by the US Air Force within its intelligence, surveillance and reconnaissance operations. She traces how the US doctrine of air-mindedness emerged and developed into a hegemonic prerogative to achieve superiority across the global sky by political, legal and technical means. The chapter by Rebecca Jones, 'From *Maschinenmensch* to Robot Bubs: Female-Presenting Autonomous Weapons Systems in Live-Action Films from 1927–2022', looks at the evolution of AWS through cinematic history with a particular focus on female representations of weapons in humanoid form. While weapons are commonly associated with male representations (with the Terminator as the most common trope), the representation of warfare is highly gendered. 'Female-presenting autonomous weapons' mirror the patriarchal gazes of their times that are merged with technical features that saliently negotiate stereotypical imaginations of the female. Jones analyses how female-presenting AWS negotiate fears and hopes of subordination, domination or (loss of) control, once more stressing how gender, power and the technical are constantly reworked with AWS. Teresa Heffernan's analysis 'Autonomous Weapons in

Fiction and the Fiction of Autonomous Weapons' also investigates the domain of fiction. She poses the question how the literal readings of fiction to animate real machines distract from the real-world development of this technology. By making reference to Karel Čapek's play *R.U.R. (Rossum's Universal Robots)* (1923) and James Cameron's *The Terminator* (1984) and its sequels, she shows how fiction has long connected the fetishization of this technology to industrial research and development. Ingvild Bode and Shimona Mohan take the reader to a completely different geographical part of the world and interrogate in 'From the Reel to the Real: Narratives of Weaponized Artificial Intelligence Technologies in India' public perspectives on AWS. Analysing survey data collected in January 2023 in India, they find that weaponized AI narratives of Anglophone countries have a high resonance among Indian respondents. At the same time, Indian respondents also share distinct ways of narrating AI technologies that integrate cultural particularities, drawing, for example, on Indian mythology and folklore as well as the mixing of genres that are typical of most Indian film productions.

Section II. Technologies and Materialities

This section looks at the concepts that are frequently applied when explaining the technological and material particularities of AWS. These include specific notions of decision making, technological agency or autonomy and debates around human-machine entanglements such as 'meaningful human control'. At the same time, the discourses on weapons technologies are always historically interwoven with the conceptual transformation of warfare and show how materialities influence particular military doctrines and vice versa:

- *ARTWORK.* «Transformator». By Peter Behrbohm, since 2013
- Lucy Suchmann: 'Il/legal War: Expanding the Frame of Meaningful Human Control from Military Operations to Democratic Governance'
- Christoph Ernst: 'From Network-Centric Warfare to Autonomous Warfighting Networks: Recontextualizing Autonomous Weapon Systems Imaginaries'
- Jens Hälterlein: 'Governing Autonomies: Imagining Responsible Artificial Intelligence in the "Future Combat Air System" European Armament Project'
- Jeremy Packer and Joshua Reeves: 'New Media, New Enemies: The Emergence of Automated Weapons in Counterterrorism'

The artwork called «Transformator» by the artist Peter Behrbohm displays the materiality and agency of an operational autonomous rocket launcher, placed in public space and policing the streets of Germany. In 'Il/legal

War: Expanding the Frame of Meaningful Human Control from Military Operations to Democratic Governance', Lucy Suchman comments on the viewpoints on the legality of AWS. She scrutinizes the debates of war that sustain militarism and how they might be challenged, not only from within but also beyond the project of arms control. She draws from her own 2016 testimony at the UN Convention on Certain Conventional Weapons (CCW), where she argued against the capacity of AWS to adhere to IHL. In her chapter she puts forward requirements of situational awareness and adherence to the principle of distinction as a necessary condition for lawful autonomy that remains unfulfilled by AWS. Christoph Ernst also points to the complicated picture of autonomy and human-machine entanglement in 'From Network-Centric Warfare to Autonomous Warfighting Networks: Recontextualizing Autonomous Weapon Systems Imaginaries'. He argues that the relevance of network-centricity for AWS imaginaries and the associated visions of future warfare is often overlooked. Ernst shows how ideas on network-centric warfare developed during the 1990s and early 2000s are the historical origins that provide important scripts and metaphors for contemporary AWS debates. By tracing this historical legacy, he argues that current AWS imaginaries contain the infrastructural vision of what can be called 'autonomous warfighting networks'. Jens Hälterlein applies these notions of networked warfare to a concrete case study in Europe. In 'Governing Autonomies: Imagining Responsible Artificial Intelligence in the "Future Combat Air System" European Armament Project', he analyses how the FCAS project imagines AI in the year 2040 as the means to enhance human decision making under the conditions of responsibility and accountability. By scrutinizing the so-called 'FCAS Ethical AI Demonstrator', he shows how FCAS applies a liberal anthropology, featuring individual responsibilization of operators and environmental management of behaviour through ethics by design – which, in his view, fails to live up to FCAS' own claims of enhancing human responsibility and accountability. The section concludes with 'New Media, New Enemies: The Emergence of Automated Weapons in Counterterrorism', in which Jeremy Packer and Joshua Reeves dive into the recursive relationship between media technology, knowledge creation and the production of political and military enemies. Through the prism of media theory, they show how media technologies produce new ways of perceiving the surrounding world and the threats that lurk therein. When applied in political or military contexts, they argue, enemies will always be uncovered, as enhanced visibility automatically brings new enemyship to the surface. They observe that with positive feedback systems, there is no way to ultimately find and neutralize all enemies. The system's operation demands the constant discovery of new problems to solve – and villains to kill.

Section III: Politics and Ethics

This section looks at the understandings and meanings of LAWS that are applied in political and ethical contexts, which are often based on 'as if' scenarios. Translated into political action, these meanings and their underlying assumptions create realities in their own right. While the actual technological capabilities are still limited, their anticipated futures nonetheless have severe implications for global security policies, regulatory and legal initiatives or military operations in light of their use by states as well as nonstate actors:

- *ARTWORK.* «XCI|XCIX, (91|99)». By Johannes Weilandt, 2023
- Elke Schwarz: 'Engineering Moral Failure? The Challenges of Algorithmic Ethics for Lethal Autonomous Weapon Systems'
- Bernhard Seidl: 'Legitimizing and Contesting Lethal Autonomous Weapons Systems in Japan: A Multilayered Analysis of Public Discourse'
- Jutta Weber: 'The Reality of (Past) Future Air Combat Systems: On Climate Wars, Carbon Costs and Rare Earth Elements'
- Thomas Christian Bächle and Xiran Liu: 'Showcasing Power, Performing Responsibility? Introducing Military Artificial Intelligence Discourses in China'

Johannes Weilandt opens the section with the artwork «XCI|XCIX, (91|99)», which shows machine-generated air images by precision and laser-guided weapons from the Second Gulf War and Yugoslav Wars of the 1990s. While now being omnipresent in times of drone and remote warfare, the images back then were broadcast on television and hailed the beginning of the era of 'smart bombs'. In 'Engineering Moral Failure? The Challenges of Algorithmic Ethics for Lethal Autonomous Weapon Systems', Elke Schwarz observes that over a decade's worth of discussions on the ethical and legal implications of AWS have yielded limited results. Problematically, these discussions are marred by unhelpful conflations, with both human agency and machine agency being read through a technological lens wherein functional equivalences are drawn between the two. She examines these discourses and their logical foundations and argues that rather than helping to make sense of the specific demand of moral agency and responsibility in the context of AWS, they take us further away from understanding moral concerns as exclusively related to humans. The political and ethical understanding of AWS remains contested – not only ethically, as Schwarz shows, but also from the viewpoint of political institutions across the globe. In 'Legitimizing and Contesting Lethal Autonomous Weapons Systems in Japan: A Multilayered Analysis of Public Discourse', Bernhard Seidl conducts an analysis of public discourse on lethal autonomous weapon systems (LAWS) in Japan. He examines texts produced in or for the public sphere, including policy documents, nongovernmental organization (NGO) material and newspapers, in order to understand how the adoption of LAWS in Japan is legitimized and contested. Seidl places his findings in the context of Japan's evolving security identity and reveals the interplay between the discourse layers and actors, realized in a language influenced by facts and imaginaries particular to the Japanese context. Evoking so much attention and allure, AWS not only have the power to attract political state interest – they also mute and sideline their hazardous side effects. In 'The Reality of (Past) Future Air Combat Systems: On Climate Wars, Carbon Costs and Rare Earth Elements', Jutta Weber discusses the carbon costs, greenhouse gas (GHG) emissions and the rare earth metal dependencies of present and future military systems. She emphasizes that the world's militaries and associated military technology industries are responsible for around 5.5 percent of global GHG emissions – without counting postwar recovery. Looking concretely at Future Combat Air Programmes and the realities of their development and deployment in the future, she argues that the emissions will ultimately inhibit the realization of these systems, rendering their future something that has already passed. Finally, in their chapter, 'Showcasing Power, Performing Responsibility? Introducing Military Artificial Intelligence Discourses in China', Thomas Christian Bächle and Xiran Liu focus on China as a major geopolitical power, which is constructed as antagonist to European and American interests. They argue that besides the actual functionality of these technologies, AI takes particular meanings that are actively utilized for government legitimacy or deterrence in political discourses. Their analysis of state media representations of military AI suggests a noteworthy gap between public agendas that promote a responsible use of AI and the actual employment of these technologies, which are used for deterrence and intimidation.

Filmography

2001: A Space Odyssey (1968) Directed by Stanley Kubrick, US: Metro-Goldwyn-Mayer.

The Birds (1963) Directed by Alfred Hitchcock, US: Alfred Hitchcock Productions.

The Terminator (1984) Directed by James Cameron, US: Hemdale, Pacific Western Productions, Euro Film Funding & Cinema '84.

References

Abu Elouf, S. (2025). Israel loosened its rules to bomb Hamas fighters, killing many more civilians. *The New York Times*. Available from: https://www.nytimes.com/2024/12/26/world/middleeast/israel-hamas-gaza-bombing.html [Accessed 05 January 2025].

Abraham, Y. (2024) '"Lavender": the AI machine directing Israel's bombing spree in Gaza', *+972 Magazine*. Available from: https://www.972mag.com/lavender-ai-israeli-army-gaza/[Accessed 12 August 2024].

Álvarez, G. (2023) 'Armas de fuego en México: pensar la violencia a partir de sus medios de ejercicio', *Este Pais*, 4 December. Available from: https://estepais.com/tendencias_y_opiniones/armas-fuego-mexico-pensando-violencia-sus-medios-ejercicio/ [Accessed 12 March 2024].

Amnesty International (n.d.) 'Arms control'. Available from: https://www.amnesty.org/en/what-we-do/arms-control/ [Accessed 11 March 2024].

Anders, G. (2002) *Die Antiquiertheit des Menschen: über die Zerstörung des Lebens im Zeitalter der dritten industriellen Revolution*, 2nd edn, Munich: CH Beck.

Antinori, A. (2019) 'Terrorism and DeepFake: from hybrid warfare to post-truth warfare in a hybrid world', in P.G. Griffiths and M.N. Kabir (eds) *Proceedings of the European Conference on Impact of Artificial Intelligence and Robotics*, Oxford: EM-Normandie Business School, pp 23–30.

Appadurai, A. (1986) *The Social Life of Things. Commodities in Cultural Perspective*, Cambridge: Cambridge University Press.

Arkin, R.C. (2009) *Governing Lethal Behaviour in Autonomous Robots*, New York: CRC Publishers.

Arquilla, J. (2021) *Bitskrieg: The New Challenge of Cyberwarfare*, Cambridge: Polity Press.

Arquilla, J. and Ronfeldt, D. (2000) *Swarming and the Future of Conflict*, Santa Monica, CA: RAND Corporation and National Defense Research Institute.

Autonomous Weapons (n.d.) 'Slaughterbots are here', *Autonomous Weapons*. Available from: https://autonomousweapons.org/ [Accessed 5 March 2024].

Barber Associates (1975) 'The Advanced Research Projects Agency, 1958–1974. II-10', in: Defense Technical Information Center. Available from: https://archive.org/details/DTIC_ADA154363 [Accessed 12 March 2024].

Bächle, T.C. (2023) 'The age of machine autonomy?', *Digital Society Blog*. https://doi.org/10.5281/zenodo.8273068

Bächle, T.C. and Bareis, J. (2022) '"Autonomous weapons" as a geopolitical signifier in a national power play: analysing AI imaginaries in Chinese and US military policies', *European Journal of Futures Research*, 10(20): 1–18. https://doi.org/10.1186/s40309-022-00202-w

Bareis, J. and Katzenbach, C. (2022) 'Talking AI into being: the narratives and imaginaries of national AI strategies and their performative politics', *Science, Technology, and Human Values*, 47(5): 855–881. https://doi.org/10.1177/01622439211030007

Bareis, J., Roßmann, M. and Bordignon, F. (2023) 'Technology hype: dealing with bold expectations and overpromising', *TATuP-Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis*, 32(3): 10–71. https://doi.org/10.14512/tatup.32.3.10

Baudrillard, J. (1995) *The Gulf War Did Not Take Place*, Bloomington: Indiana University Press.

Bennett, D. and Ilyushina, M. (2023) 'Inside the Russian effort to build 6,000 attack drones with Iran's help', *Washington Post*, 17 August. Available from: https://www.washingtonpost.com/investigations/2023/08/17/russia-iran-drone-shahed-alabuga/ [Accessed 11 March 2024].

Beckert, J. (2016) *Imagined Futures*, Cambridge, MA: Harvard University Press.

Bhuta, N., Beck, S., Geiß, R., Liu H.Y. and Kreß, C. (2016) *Autonomous Weapons Systems: Law, Ethics, Policy*, Cambridge: Cambridge University Press.

Block, F. (2008) 'Swimming against the current: the rise of a hidden developmental state in the United States', *Politics & Society*, 36(2): 169–206. https://doi.org/10.1177/0032329208318731

Bousquet, A. (2018). *The Eye of War: Military Perception from the Telescope to the Drone*. Minneapolis: University of Minnesota Press.

Bousquet, A. (2022). *The Scientific Way of Warfare: Order and Chaos on the Battlefields of Modernity*, Oxford: Oxford University Press.

Bousquet, A., Grove, J. and Shah, N. (2020) 'Becoming war: towards a martial empiricism', *Security Dialogue*, 51(2–3): 99–118. https://doi.org/10.1177/0967010619895660

Bruno, M. (2022) '"Uber for artillery": What is Ukraine's GIS Arta System?', *The Moloch*, 24 August. Available from: https://themoloch.com/conflict/uber-for-artillery-what-is-ukraines-gis-arta-system/ [Accessed 11 March 2024].

Business Research Company (2024) 'Autonomous military weapons global market report'. Available from: https://thebusinessresearchcompany.com/report/autonomous-military-weapons-global-market-report [Accessed 12 August 2024].

Campolo, A. and Crawford, K. (2020) 'Enchanted determinism: power without responsibility in artificial intelligence', *Engaging Science, Technology, and Society*, 6: 1–19. https://doi.org/10.17351/ests2020.277

Čapek, K. (1923) 'R.U.R. (Rossum's Universal Robots)'. Available from: https://www.gutenberg.org/cache/epub/59112/pg59112-images.html [Accessed 12 August 2024].

Cave, S. and Dihal, K. (2019) 'Hopes and fears for intelligent machines in fiction and reality', *Nature Machine Intelligence*, 1(2): 74–78. https://doi.org/10.1038/s42256-019-0020-9

Cayirci, E., AlNaimi, R. and AlNabet, S. S. (2022) 'Computer assisted military experimentations', in *2022 Winter Simulation Conference (WSC)*, Singapore, pp 1311–1324. http://dx.doi.org/10.1109/wsc57314.2022.10015294

CCW Group of Governmental Experts on LAWS (2018) 'Position paper', *Reaching Critical Will*, April. Available from: https://www.reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2018/gge/documents/GGE.1-WP7.pdf [Accessed 7 March 2024].

Cebrowski, A.K. (2005) 'The implementation of network-centric warfare', *Office of Force Transformation Washington*. Available from: https://apps.dtic.mil/sti/citations/ADA446831 [Accessed 7 March 2024].

Coeckelbergh, M. (2011) 'From killer machines to doctrines and swarms, or why ethics of military robotics is not (necessarily) about robots', *Philosophy & Technology*, 24(3): 269–278. https://doi.org/10.1007/s13347-011-0019-6

Cohen, E.A. (1996) 'A revolution in warfare', *Foreign Affairs*, 75(2): 37–54. https://doi.org/10.2307/20047487

Cummings, M.L. (2017): 'Artificial intelligence and the future of warfare'. Available from: https://www.chathamhouse.org/sites/default/files/publications/research/2017-01-26-artificial-intelligence-future-warfare-cummings-final.pdf, last visited 2020-20-01 [Accessed 7 March 2024].

DARPA (n.d.) 'Homepage'. Available from: https://www.darpa.mil/ [Accessed 12 March 2024].

DARPAtv (2018) 'Darpa overview', *YouTube*, 16 January. Available from: https://www.youtube.com/watch?v=IcV4pgB5QY0 [Accessed 12 March 2024].

Davies, H., McKernan, B. and Sabbagh, D. (2023) '"The Gospel": how Israel uses AI to select bombing targets in Gaza', *The Guardian*, 1 December. Available from: https://www.theguardian.com/world/2023/dec/01/the-gospel-how-israel-uses-ai-to-select-bombing-targets [Accessed 5 March 2024].

De Boisboissel, G. and Le Masson, J.-M. (2021) 'The enhanced soldier: definitions', *Military Review*: 1–5. Available from: https://www.armyupress.army.mil/Journals/Military-Review/Online-Exclusive/2021-French-OLE/Part-2-Definitions/ [Accessed 12 March 2024].

De Landa, M. (1991) *War in the Age of Intelligent Machines*, New York: Zone Books.

Diamond, J.M. (1997) *Guns, Germs, and Steel: The Fates of Human Societies*, New York: W.W. Norton.

Echevarria, A.J. and Shaw, J.M. (1992) 'The new military revolution: post-industrial change', *Parameters Winter*, 22(1): 70–79. https://doi.org/10.55540/0031-1723.1624

The Economist (2024) 'How Ukraine is using AI to fight Russia', 8 April. Available from: https://www.economist.com/science-and-technology/2024/04/08/how-ukraine-is-using-ai-to-fight-russia [Accessed 12 August 2024].

Elish, M.C. (2018) 24/7: *Drone Operations and the Distributed Work of War*, New York: Columbia University in the City of New York.

Favaro, M. and Schwarz, E. (2022) 'Human augmentation and nuclear risk: the value of a few seconds', *Arms Control Today*, 54.

Galison, P. (1994) 'The ontology of the enemy: Norbert Wiener and the cybergenetic vision', *Critical Inquiry*, 21(1): 228–266.

GIS ARTA (n.d.) 'GIS "ARTA": automated command and control system'. Available from: https://gisarta.org/en/index.html [Accessed 5 March 2024].

Goecks, V.G. and Waytowich, N. (2024) *COA-GPT: Generative Pre-trained Transformers for Accelerated Course of Action Development in Military Operations*, Koblenz: NATO Science and Technology Organization Symposium (ICMCIS) by the Information Systems Technology (IST) Panel. https://doi.org/10.48550/arXiv.2402.01786

Graae, A.I. and Maurer, K. (2021) *Drone Imaginaries: The Power of Remote Vision*, Manchester: Manchester University Press.

Gramlich, A. (2023) 'What the data says about gun deaths in the U.S.', *Pew Research Center*, 26 April. Available from: https://www.pewresearch.org/short-reads/2023/04/26/what-the-data-says-about-gun-deaths-in-the-u-s/ [Accessed 11 March 2024].

Grand View Research (n.d.) *Artificial Intelligence In Military Market Size, Share & Trends Analysis Report By Offering, By Application, By Technology, By Platform, By Installation, By Region, And Segment Forecasts, 2023–2030*. Available from: https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-military-market-report# [Accessed 12 March 2024].

Gun Violence Archive (2024) *Past Summary Ledgers*. Available from: https://www.gunviolencearchive.org/past-tolls [Accessed 5 January 2025].

Haraway, D. (2006) 'A cyborg manifesto', in S. Stryker and S. Whittle (eds) *The Transgender Studies Reader*, New York: Routledge, pp 103–118.

Horwitz, R.B. (2022) 'Trump and the "deep state"', in T.S. James (ed) *The Trump Administration*, New York: Routledge, pp 41–58.

Hoskins, A. and O'Loughlin, B. (2015) 'Arrested war: the third phase of mediatization', *Information, Communication and Society*, 18(11): 1320–338. https://doi.org/10.1080/1369118X.2015.1068350

Hughes, J. (2004) 'Deconstructing the bomb: recent perspectives on nuclear history', *British Journal for the History of Science*, 37(4): 455–464. https://doi.org/10.1017/S0007087404006168

Jasanoff, S. (2015) 'Future imperfect: science, technology, and the imaginations of modernity', in S. Jasanoff and S.-H. Kim (eds) *Dreamscapes of Modernity*, Chicago: University of Chicago Press, pp 1–33.

Jasanoff, S. and Kim, S.-H. (2009) 'Containing the atom: sociotechnical Imaginaries and nuclear power in the United States and South Korea', *Minerva*, 47: 119–146. https://doi.org/10.1007/s11024-009-9124-4

Jasanoff, S. and Kim, S.-H. (2015) *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*, Chicago: University of Chicago Press.

JEDI Foundation (n.d.) 'Homepage'. Available from: https://www.jedi.foundation/ [Accessed 5 March 2024].

Johnson, J. (2020) 'Deterrence in the age of artificial intelligence and autonomy: a paradigm shift in nuclear deterrence theory and practice?', *Defense and Security Analysis*, 36(4): 422–448. https://doi.org/10.1080/14751798.2020.1857911

Karppi, T., Böhlen, M. and Granata, Y. (2018) 'Killer robots as cultural techniques', *International Journal of Cultural Studies*, 21(2): 107–123. https://doi.org/10.1177/1367877916671425

Khumalo, T. (2023) 'South Africa: spike in gun crime angers citizens', *Deutsche Welle*, 7 March. Available from: https://www.dw.com/en/south-africa-spike-in-gun-crime-angers-citizens/a-64903654[Accessed 12 March 2024].

Kirsten, A. (2023) 'Guns and gun violence in South Africa – the evidence', *Daily Maverick*, 16 February. Available from: https://www.dailymaverick.co.za/article/2023-02-16-guns-and-gun-violence-in-south-africa-the-evidence/ [Accessed 12 March 2024].

Krasmann, S. and Weber, J. (2015) 'Game changer? On the epistemology, ontology and politics of drones', *Behemoth: A Journal on Civilisation*, 8(2): 3–11. https://doi.org/10.6094/behemoth.2015.8.2.866

Krishnan, A. (2016) *Killer Robots: Legality and Ethicality of Autonomous Weapons*, New York: Routledge

Latour, B. (1993) *We Have Never Been Modern*, Cambridge, MA: Harvard University Press.

Lenoir, T. and Caldwell, L. (2018) *The Military-Entertainment Complex*, Cambridge, MA: Harvard University Press.

Lin, P., Bekey, G. and Abney, K. (2008) 'Autonomous military robotics: risk, ethics, and design'. *Cal Poly*, 20 December. Available from: https://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1001&context=phil_fac [Accessed 7 March 2024].

Lubell, N. (2023) 'Keynote lecture: AI in critical decision: is human judgment a legal requirement?', *DILEMA 2023 Conference*, 12 October. Available from: https://www.asser.nl/media/796679/dilema-2023-conference-programme.pdf [Accessed 13 March 2024].

Lushenko, P. and Carter, K. (2024) 'A new military-industrial complex: how tech bros are hyping AI's role in war', *Bulletin of the Atomic Scientists*. Available from: https://thebulletin.org/2024/10/a-new-military-industrial-complex-how-tech-bros-are-hyping-ais-role-in-war/October [Accessed 15 October 2024].

Lyon, D. (2015) *Surveillance after Snowden*, Hoboken, NJ: John Wiley & Sons.

Marx, L. (2000) *The Machine in the Garden: Technology and the Pastoral Ideal in America*, New York: Oxford University Press.

Mazzucato, M. (2011) 'The entrepreneurial state', *Soundings*, 49(49): 131–142. https://doi.org/10.3898/136266211798411183

Merrin, W. (2018) *Digital War: A Critical Introduction*, New York: Routledge.

Natale, S. and Ballatore, A. (2020) 'Imagining the thinking machine: technological myths and the rise of artificial intelligence', *Convergence*, 26(1), 3–18. https://doi.org/10.1177/1354856517715164

Noorman, M. and Johnson, D.G. (2014) 'Negotiating autonomy and responsibility in military robots', *Ethics and Information Technology*, 16: 51–62. https://doi.org/10.1007/s10676-013-9335-0

NSSF (2022) 'Firearm and ammunition industry: economic Impact'. Available from: https://www.nssf.org/wp-content/uploads/2022/03/2022-Firearm-Ammunition-Industry-Economic-Impact.pdf [Accessed 12 March 2024].

Nye, D.E. (2004) 'Technological prediction: a Promethean problem', in M. Sturken, D. Thomas and S. Ball-Rokeach (eds) *Technological Visions: The Hopes and Fears That Shape New Technologies*, Philadelphia, PA: Temple University Press, pp 159–176.

Packer, J. and Reeves, J. (2020). *Killer Apps: War, Media, Machine*, Durham, NC: Duke University Press.

Permanent Representation of the Federal Republic of Germany to the Conference on Disarmament in Geneva (2018) 'Statement delivered by Germany on working definitions of LAWS/"definition of systems under consideration"', *Reaching Critical Will*, 9 April. Available from: https://reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2018/gge/statements/9April_Germany.pdf [Accessed 7 March 2024].

Rarm, L. (2023) 'Terror: *live*', *Continuum*, 37(3): 422–432. https://doi.org/10.1080/10304312.2023.2255395

Rasmussen, N. (2011) 'Medical science and the military: the Allies' use of amphetamine during World War II', *Journal of Interdisciplinary History*, 42(2): 205–233. https://doi.org/10.1162/JINH_a_00212

The Red Team (2021) 'Discover the Red Team'. Available from: https://redteamdefense.org/en/meet-the-red-team [Accessed 12 March 2024].

Richardson. M, (2020) 'Drone cultures: encounters with everyday militarisms', *Continuum*, 34(6): 858–69. https://doi.org/10.1080/10304312.2020.1842125

Richardson, M. (2024) *Nonhuman Witnessing: War, Data, and Ecology after the End of the World*, Durham, NC: Duke University Press.

Rupka, S. and Baggiarini, B. (2018) 'The (non) event of state terror: drones and divine violence', *Critical Studies on Terrorism*, 11(2): 342–356. https://doi.org/10.1080/17539153.2018.1456735

Rudloff, J.P. amd Appel, M. (2023) 'Fake news', in M. Appel, F. Hutmacher, C. Mengelkamp, J.P. Stein and S. Weber (eds) *Digital ist besser?! Psychologie der Online- und Mobilkommunikation*, Berlin: Springer, pp 217–232.

Sarasin, P. (2006) *Anthrax: Bioterror as Fact and Fantasy*, Cambridge, MA: Harvard University Press.

Sauer, F. (2016) 'Stopping "killer robots": why now is the time to ban autonomous weapons systems', *Arms Control Today*, 46(8).

Scharre, P. (2018) *Army of None: Autonomous Weapons and the Future of War*, New York: W.W. Norton and Company.

Schwarz, E. (2018). *Death Machines: The Ethics of Violent Technologies*, Manchester: Manchester University Press.

Seib, P. (2021) *Information at War: Journalism, Disinformation, and Modern Warfare*, Cambridge: Polity Press.

Shaw, I.G.R. (2016) *Predator Empire: Drone Warfare and Full Spectrum Dominance*, Minneapolis: University of Minnesota Press.

Singer, P.W. (2010) *Wired for War: The Robotics Revolution and Conflict in the Twenty-First Century*, New York: Penguin Books.

Simonite, T. (2021) '3 years after the Project Maven uproar, Google cozies to the Pentagon', *WIRED*, 10 November. Available from: https://www.wired.com/story/3-years-maven-uproar-google-warms-pentagon/ [Accessed 12 March 2024].

Stop Killer Robots (n.d.) 'Homepage'. Available from: https://www.stopkillerrobots.org/ [Accessed 5 March 2024].

Stoner, R.H. (2009) 'R2D2 with attitude: the story of the phalanx close-in weapons', *NavWeaps*, 30 October. Available from: http://www.navweaps.com/index_tech/tech-103.php [Accessed 7 March 2024].

Suchman, L. (2023) 'The uncontroversial "thingness" of AI', *Big Data & Society*, 10(2): 1–5. https://doi.org/10.1177/20539517231206794

Suchman, L. and Weber, J. (2016) 'Human-machine autonomies', paper presented at the Autonomous Weapons Systems – Law, Ethics, Policy Conference, 24–25 April 2014, Academy of European Law and European University Institute, Florence. Available from: https://www.researchgate.net/publication/270104965_Human-Machine_Autonomies [Accessed 10 January 2025].

Toffler, A. and Toffler, H. (1993) *War and Anti-war: Making Sense of Today's Global Chaos*, New York: Warner Books.

UK Ministry of Defence/Bundeswehr Office for Defence Planning (2021) *Human Augmentation: The Dawn of a New Paradigm*, London: UK Ministry of Defence.

US Department of Defense (2023) 'DoD Directive 3000.09 autonomy on weapon systems', *Washington Headquarters Service*, 25 January. Available from: https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf [Accessed 13 March 2024].

Von Clausewitz, C. (1942) 'Principles of war', *Clausewitz Studies*. Available from: https://www.clausewitzstudies.org/mobile/principlesofwar.htm [Accessed 12 March 2024].

Williams, A. (2015) 'Defining autonomy in systems: challenges and solutions', in A. Williams and P. Scharre (eds) *Autonomous Systems. Issue for Defence Policymakers*, Norfolk, VA: NATO Headquarters Supreme Allied Commander.Transformation, pp 27–62.

Worcester, M. (2015) 'Autonomous warfare: a revolution in military affairs', *ISPSW Strategy Series: Focus on Defense and International Security*, 340: 1–6.

Work, R.O. and Brimley, S. (2014) *20YY: Preparing for War in the Robotic Age*, Washington DC: Center for a New American Century.

Zikusoka, D. (2023) 'How Ukraine's "Uber for artillery" is leading the software war against Russia', *New America*, 25 May. Available from: https://www.newamerica.org/future-frontlines/blogs/how-ukraines-uber-for-artillery-is-leading-the-software-war-against-russia/ [Accessed 11 March 2024].

I

Narratives and Theories

ARTWORK

The Unreachable Myth: Killing Unknown Victims with Insensible Means by Unidentified Perpetrators for Unapparent Reasons

Jinyu Wang, 2023

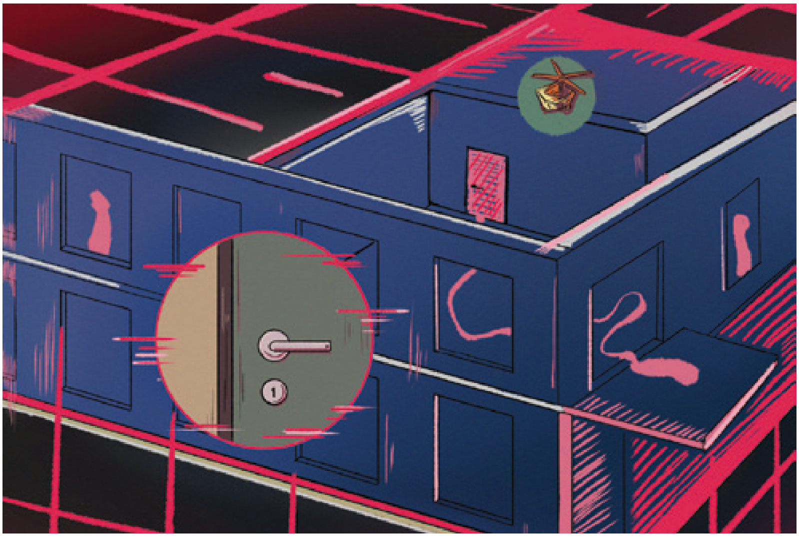
Artificial intelligence (AI) has been rapidly building a new type of railroad tracks around the world. This time, humankind will have to accept not only iron horses, but also the fact that they themselves are being made of iron. The loud roar of the wheels rolling over the railroad tracks once kept people awake at night. Now, is it the enemy or the self that AI is aiming at with its raising gun?

The illustrations displayed here are part of an artwork series that I produced as a student at HBK Saar (Saar College of Fine Arts) in Saarbrücken, Germany under the supervision of Mert Akbal. They depict the functions and the impact of military AI in wars and combat, combining the reality of military AI's operations with its influence on future landscapes. Aimed to interrogate the functionality of military AI, each illustration contains a story based on different effects and the predictable future humans will face.

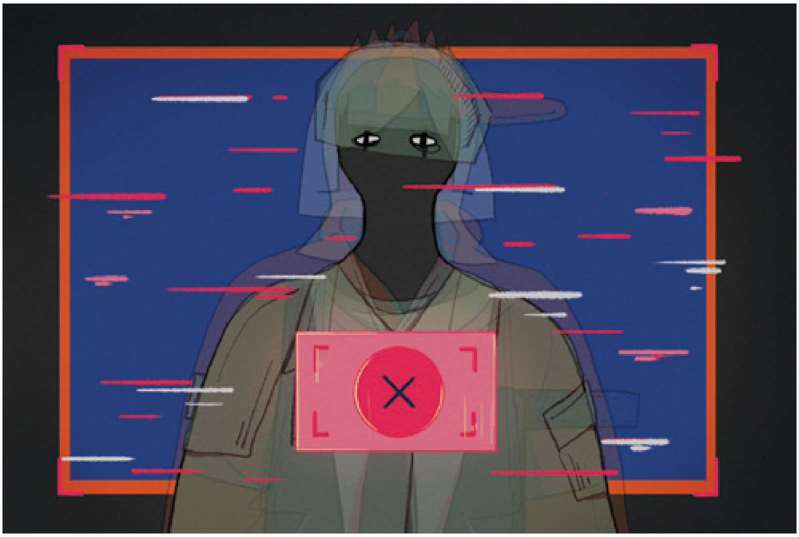
Inspired by sci-fi art from the 1970s, they are illustrated with a relatively neutral and flat style, making the topic more accessible without losing sight of their technological reality. This collection tries to help explain and expand discussions about autonomous weapon systems (AWS) in political and academic discourses, they carefully show how real-life situations and possible future scenarios are connected. By combining facts in different academic fields, they expand the borders of each subject and provide new perspectives on this sensitive topic.

The collection demonstrates how autonomous weapons kill under the nonrecognition of their targets, their operators and their reasons.

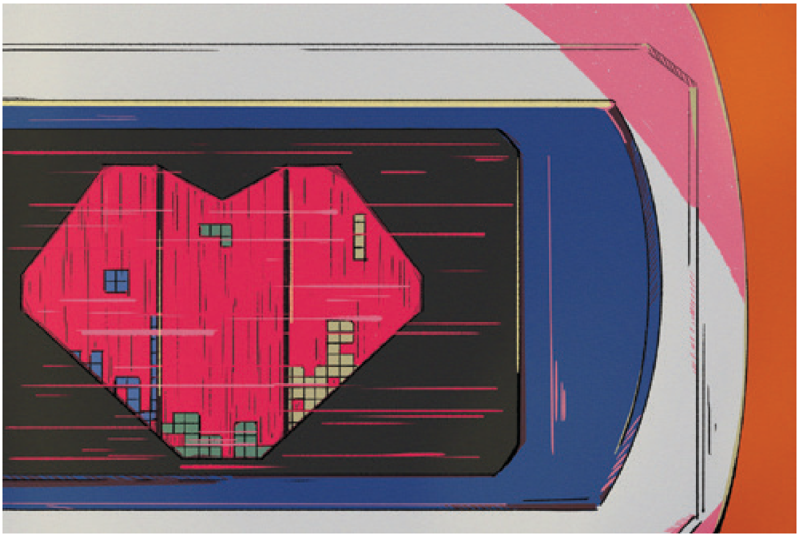
The first illustration shows the virtual image military AIs map, when they are assigned with and executing killing missions. Military AIs cannot understand who exactly it is they are killing, what kind of personality they have, or the reason behind their current actions. The second illustration deals with how AWS receive data, analyse information and recognize targets. Moving to the third illustration, a Tetris game stands in the middle of a robot screen, without any identifiable players, as if there were no perpetrators who assign those killing missions: Military AI will take full responsibility as a result. Nevertheless, nobody will deny that on each military AI, there will be a clear symbol showing where they belong to – and who earns better scores around the world in the end. The fourth illustration questions the neutrality about military AI, with an attempt to figure out the killing criteria with which military AI aligns. Massive amounts of data and algorithms have hidden the motives, as machines and AIs tend to be seen as neutral and rational agents, obfuscating the fact that data itself can be originally discriminatory and unjust. Facing the illusion of justice, it will be more difficult to see the reasons behind AI's killing action.



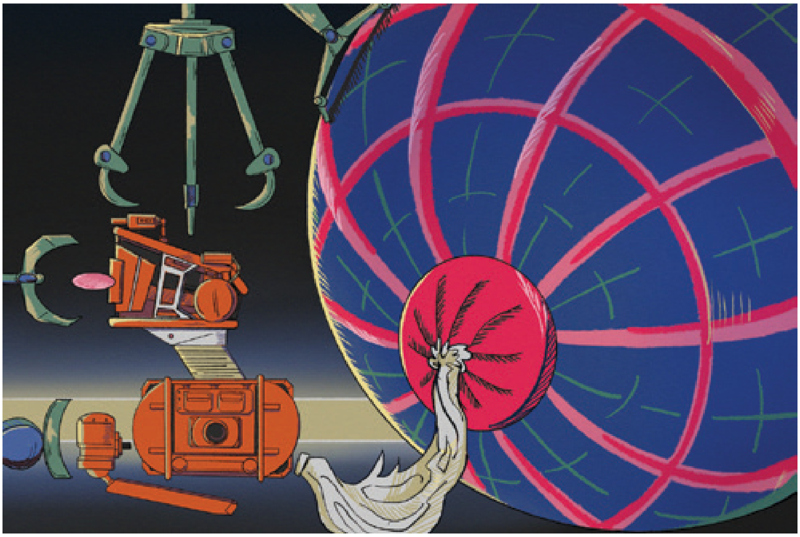
1. *Unknown techniques*



2. *Unrecognisable victims*



3. *Unidentifiable perpetrators*



4. *Unapparent motives*

The AI/lure of US Airpower: Imaginaries of Disruption in the Pursuit of Technological Superiority Since the Early 20th Century

Jennifer Rooke

Introduction: disruptive technologies

> Today many modern war devices of great destructive power can be built piecemeal and under cover. Sub-assemblies might be secretly made in underground laboratories, and assembled into an annihilating war machine. War may descend upon us by thousands of robots passing unannounced across our shorelines – unless we act now to prevent them. Today, Japanese and German cities lie in ruins, but they merely suggest the vast destruction that can be done with the weapons of tomorrow. The first target of a potential aggressor might well be our industrial system or our major centers of population. If the United States is to be secure in the future, we must never relinquish the means of preventing such a blow. (Arnold, 1945: 60)

A few months after Japan's statement of surrender that brought an official end to the Second World War, Henry 'Hap' Arnold, the Commanding General of the United States Army Air Forces (USAAF) – the direct organizational predecessor to today's United States Air Force (USAF)[1] – submitted the final

[1] The lineage of the US Air Force remained subordinate to the Army across six different organizational transformations until it achieved independence through the National Security Act of 1947: US Army Signal Corps Aeronautical Division, 1907–1914; US

in a series of three reports to the War Secretary in which he highlighted perceived successes targeting supposed vital enemy installations and logistical flows. He included photos of damage to aircraft production facilities caused by a series of incendiary bombing attacks across Japan's industrial cities and candidly stated that US B-29 bomber aircraft made the country 'bleed internally' (Arnold, 1945: 36). Missing from those photos was the devastating violence inflicted upon the victims of what have been documented as the deadliest airstrikes in history (an estimated combined 187,000 deaths and 214,000 injuries, with approximately 100,000 of those deaths inflicted during a single night over Tokyo on 9–10 March 1945),[2] deadlier even than the two atomic bombs dropped over Hiroshima (an estimated 80,000 instant deaths) and Nagasaki (an estimated 40,000 instant deaths). In one of his final initiatives as the last USAAF commanding general, Arnold proceeded to outline a need for a separate Air Force on equal standing with the Army and Navy: 'Air superiority accordingly is the first essential for effective offense as well as defense. A modern, autonomous, and thoroughly trained Air Force in being at all times will not alone be sufficient, but without it there can be no national security' (Arnold, 1945: 59). He also advanced the argument for continued collaboration with industry in the face of significant demobilization in wartime production that had enabled accelerated experimentation with and production of new aerial weapon systems, epitomized by the Manhattan Project that resulted in the atomic bomb and spurred a nuclear missile arms race: 'Since military Air Power depends for its existence upon the aviation industry and the *air-mindedness* of the nation, the Air Force must promote the development of American civil Air Power in all of its forms, both commercial and private' (Arnold, 1945: 70, emphasis added).

This concept of air-mindedness, and the violence it engenders, is the focus of this chapter: an analysis of the sociotechnical imaginaries that shape – and are shaped by – today's use of lethal autonomous weapon systems (AWS)-related technologies within the USAF, and specifically within its intelligence, surveillance and reconnaissance (ISR) operations. This chapter proceeds from a historical perspective in order to demonstrate the long continuity of these imaginaries. It analyses a political rationality that can be traced to the discourse and practices surrounding the advent of the airplane in warfare from which the USAF eventually emerged

Army Signal Corps Aviation Section, 1914–1918; US Army Air Service, 1918–1926; US Army Air Corps, 1926–1941; and, US Army Air Forces, 1941–1947.

[2] For additional information on these airstrikes, including artwork by survivors, see the bilingual English-Japanese digital archive available from: japanairraids.org

as an independent department within the US military to gain political influence on a par with that of the Army and the Navy. The airplane and the atomic bomb have been variously characterized as *revolutionary* or *war-changing* technologies. AWS technologies are similarly described, but are more commonly coined by the US military as *disruptive* technologies. The USAF announced in 2018 through a publicly released document and video that it had produced a ten-year *Next Generation ISR Dominance Flight Plan* as an investment strategy framework to meet challenges associated with 'the changing character of war based on disruptive technologies' (Mitchell Institute, 2018: 09:15), further describing them as 'neural networks, deep learning, human-machine teaming, and artificial intelligence' (AI) that industry experts encompass within the lexicon of 'machine intelligence' (Mitchell Institute, 2018: 10:45). This chapter explores the logics, underlying assumptions and human judgements prompting this perception of disruptiveness that emanates from anxiety – a moral panic even – about a potential loss of dominance and perceived superiority. The advent of these technologies is thus viewed as a threat, but also as an opportunity – one propelling the US towards an annual trillion dollar defence budget in order to maintain a so-called competitive advantage.

This chapter describes a community in which I was socialized for almost 30 years, from 1985 as a university freshman cadet at the US Air Force Academy until military retirement in 2013 at the rank of colonel. I served as an intelligence officer and over the course of that career heeded a rationality that embraces a racialized and ideologically motivated violence inflicted with the aim of exporting a universalizing version of democracy and so-called 'development' throughout the world. It is a version of democracy based on a sense of moral superiority with presumed authority to determine who (and under what conditions) gets included or excluded in 'the modern international' (Walker, 2006: 58). It is a political rationality that decides which lives are worth protecting and saving, and which are expendable. I have come to question the underlying presuppositions of this political rationality, and this chapter aims to highlight the contradictions it entails. In the first section on air-mindedness of the nation, it offers an account of early airpower advocacy focused on a specific vision of aerial warfare – strategic paralysis – that has shaped enduring technical, political and legal means to achieve air superiority and that remain operative within the USAF's contemporary visions of information warfare that employ emerging AWS technologies to achieve information superiority. In the second section on imminent threats, it then proceeds to unpack the underlying presuppositions that define certain entities as existential threats to be eliminated. Next, in the third section on identifying the enemy, it highlights the significant degree of uncertainty inherent in ISR operations. Finally, in the conclusion section on meaningful human control, it offers insights relating to regulating the use of lethal AWS technologies.

1. The air-mindedness of the nation: an acculturation to US airpower over the course of the 20th century

> Tomorrow, new generations will reconstruct what we are destroying: but we must not fear to destroy if by so doing we may have victory, nor to kill, if by death only we may have life … Let the deadly rain borne by Italian wings, fall from the skies! No one will ever condemn us for having killed the war! (Salvaneschi, 1917: 66–68)

> [Do] the people who live in our village have the right to a decent life and to live without the fear of being killed by a drone? (Khaled Mohmed Naser bin Ali Jaber, Yemen drone attack survivor, as quoted in Turse, 2021)

This section traces some of the imaginaries that have driven expansions of US military airpower and engendered an air-mindedness among US government officials, as well as the public that to a considerable degree have come to unquestioningly accept the necessity of offensive air strikes abroad in pursuance of an ostensible national self-defence, both within and beyond officially designated warzones. Instead of devising touted means to end war quickly, early US military airpower advocates – active in different phases between 1917 during the First World War and 1945 at the end of the Second World War – simply extended its horrors above and beyond the battlespaces of military combatants. They targeted economic centres and civilian populations deep into enemy territory beyond traditional tactical lines of battle, a doctrine referred to as strategic bombing. Two key concepts within military doctrine – proportionality and precision – also arose during this timeframe that directly pertain to contemporary interpretations of customary international humanitarian law (IHL). After the Second World War, new iterations of aerial warfare strategy and doctrine were spurred on by a combination of activities: the IHL treaties composing the Geneva Conventions of 1949 that the US ratified in 1955, the ineffectiveness of and negative reactions to strategic bombing that continued during the US wars in Korea (1950–1953) and Vietnam (1964–1973), and developments in ISR collection technology that helped refine the target selection process. This phase of airpower theorizing, referred to as strategic paralysis, moved away from targeting enemy economic centres in large metropolitan areas to focus on paralysis of the ability of an enemy's leadership to control and replenish their fighting forces. The concept of strategic paralysis continues to underpin current USAF operations. The family members[3] of Khaled Mohmed Naser bin Ali Jaber, among countless other noncombatant victims of US drone attacks, were targeted in 2012 outside their village mosque in Yemen because they were erroneously perceived as playing some role in the sustainment of the fighting forces of purported terrorist networks (Turse, 2021). Bin Ali Jaber's challenge in the form of a question to the US president, which is quoted at the beginning of this section, was rhetorical because he knew the answer: he does not need permission in the form of a right to exist bestowed upon him by someone else. It was a claim to force a response, a recognition of his humanity and dignity.

1.1 European colonial imaginaries and the first aerial bombing attack (1907–1911)

The invention of the airplane sparked imagination, especially across Europe and the US, and depicted particularly by European novelists such as Rudolf Martin (*Berlin-Bagdad*, 1907), H.G. Wells (*The War in the Air*, 1908), Emile Driant (*L'Aviateur du Pacifique; The Aviator of the Pacific*, 1910 and *Au-dessus du Continent Noir; Above the Dark Continent*, 1911), and Gabriele d'Annunzio (*Forse che sì forse che no; Maybe Yes, Maybe No*, 1910), who rapidly popularized visionary ideas about the future of war from the air even before militaries seriously considered the airplane as a viable technology for waging war. Historian Robert Wohl has analysed these narratives as reflections of the prevailing nationalistic sentiments circulating within each author's respective country at that time. They portrayed future air warfare across a continuum of dystopian-utopian imaginaries. They all generated imagery of expansion and decline and of a hierarchy based on those who possessed this new technology and the knowhow to harness its potential. H.G. Wells, for example, conjured up a German air fleet that attacked New York City without warning: 'Below, they left ruins and blazing configurations and heaped and scattered dead: men, women, and children mixed together as though they had been no more than Moors, or Zulus, or Chinese' (Wohl, 1994: 74). He also evoked a mysterious world beyond 'Christendom' led by an 'Asiatic Confederation', a 'yellow peril' that laboured in hidden factories to devise superior aerial weapons with which to deliver a fatal blow to their enemies (Wohl, 1994: 91) – a tale

[3] Most victims of US airstrikes only become identified as one number among many that are roughly estimated and counted. Many of them remain uncounted. The story of Salem bin Ali Jaber, a cleric from the village of Kashamir in Yemen, and his cousin Waleed bin Ali Jaber, a local policeman, can be read in Nick Turse's article (Turse, 2021).

similarly retold almost 40 years later in General Arnold's *Third Report*, quoted at the beginning of this chapter. Though the language used to characterize this perceived threat has changed, such sentiments remain discernible in 21st century US Department of Defense (DoD) documents, such as the annual *Report on Military and Security Developments Involving the People's Republic of China*. Driant portrayed scenes of awe-struck Africans too astounded by the spectacle of flying machines to shoot back (see Figures 1 and 2), thus recognizing the overwhelming superiority of their 'white masters' (Wohl, 1994: 89) – sentiments echoed more than

Figure 1: Cover of Emil Driant's 1911 book *Au-Dessus du Continent Noir*

Figure 2: 'The formidable machine, in a second, sowed death' (Driant, 1911: 433, my translation from original in French)

80 years later by the US military to describe its paralysing 'shock and awe' strategies of dominance in Iraq.

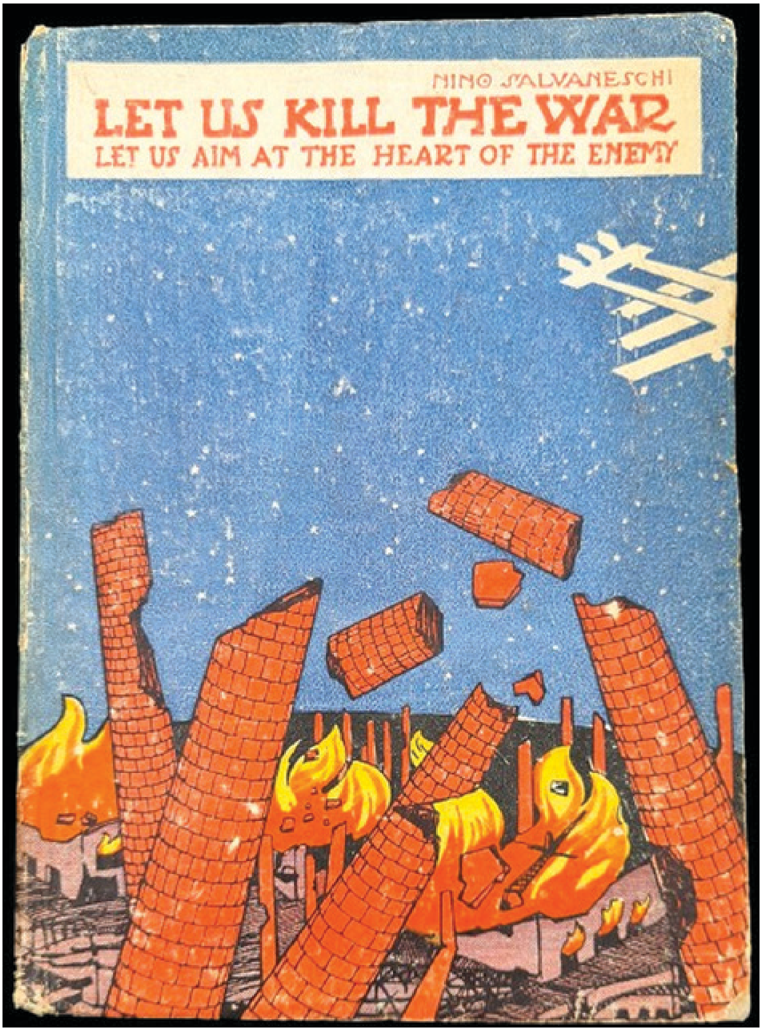
In Italy, d'Annunzio shared many of these same sentiments with the avant-garde art community and its emerging Futurism Movement that flourished between 1909 and 1944. Through their manifestos and artwork, they professed an ethos of dynamism and celebrated speed, technology, and violence. These ideas were also embraced by Guilio Douhet, a military officer recognized as one of the earliest, most vocal airpower theorists. His work and writing would in turn directly influence US air war planning once the US entered the First World War. In 1909, Douhet extended his ideas beyond popular fiction and art, and turned them into a military treatise (Gat, 1997). He surmised that warfare in the third dimension of space above the ground and sea would become a crucial battlefield in which those who could control the air would also control those surfaces; that aircraft would transport war beyond traditional battle lines to ‘all people in all places’; and that the ‘psychological effects of aerial bombardment would be great’ (Meilinger, 1997a: xiii–iv). Less than two years later, the Kingdom of Italy invaded Ottoman Tripolitania. During that colonial conquest, the Italians flew the first-ever reconnaissance mission from an airplane, above the Ain Zara oasis near Tripoli, that was followed nine days later by the first-ever aerial bomb attack. The pilot, described in Italian press at the time as a ‘flying artilleryman’ who invented the ‘art of winged death’, expected to find the camp of about 2,000 Arab and Turkish fighting forces over which to drop his grenades (Maksel, 2011). The Ottomans claimed that they fell on a field hospital (Patterson, 2011).

Contestation over the violence that aerial warfare inflicted thus arose at its inception. However, it would take more than a decade after the first aerial bomb attack in 1911 before an effort to regulate the indiscriminate weaponry of aerial warfare resulted in the drafting in 1923 of *The Hague Air Rules* that ultimately failed ratification because the states that used those weapon systems did not want to give up what they considered decisive, ‘more humane’ weapons of ‘peace’ (van Dijk, 2022: 209–210). These draft rules outlined which types of objects could be targeted and which should be prohibited *proportionally* through a calculation of military importance in relation to the danger posed to civilian populations, thus maintaining the primacy of strategic bombing interests over the protection of civilians (van Dijk, 2022: 209). Populations deemed expendable would endure the devastation of incendiary fire and atomic bombing before that language would be officially incorporated into IHL after the Second World War.

1.2 Early airpower imaginaries to terrorize civilians: the inception of US strategic bombing plans (1917–1918)

Most scholars who attribute Douhet’s influence on US airpower theorizing and practice situate it after the First World War in the 1920s based on access to an English translation of his postwar book *Il Dominio dell’aria*; *The Command of the Air*, published in 1921. However, a direct link was made earlier in 1917, shortly after the US officially entered the war as a member of the Entente Powers, commonly referred to as the Allies. A US military delegation travelled to Europe to select which allied aircraft would be most suitable for construction at home. One team member, Major Edgar Gorrell, met with industrialist

Figure 3: Cover of Nino Salvaneschi's 1917 pamphlet



Giovanni Caproni, who, with Douhet's urging and support, had developed one of the world's first long-range bombers (Gat, 1997; Meilinger, 1997b). Gorrell requested information on German industrial targets that the US could use in its mission planning. Caproni delivered a document containing Douhet's plans and included a polemic penned by Nino Salvaneschi, a Futurist and friend of Douhet. Gorrell distributed copies that proclaimed '*Let Us Kill the War, Let Us Aim at the Heart of the Enemy!*' (see Figure 3) within the American Expeditionary Force Air Service headquarters in France that was led briefly by Colonel William 'Billy' Mitchell, who would later become mythologized by his ardent supporters as the 'father' of the USAF (Meilinger, 1997b: 6–7).

Gorrell directed efforts that would produce the first US air campaign plan to drop bombs deep into enemy territory beyond the line of battle trenches in order to disrupt their links to logistical supplies and industrial production. The US never implemented that plan due to production and shipping problems that hampered the deployment of bomber aircraft to Europe until only two days before the armistice, so their aerial contributions remained restricted to pursuit and observation aircraft missions (Clodfelter 1997: 85–88). The British, French, Germans, Italians and Russians had all rapidly developed and experimented with employment of bomber aircraft throughout the First World War, frightening urban populations below – from London, Liège and Paris to Freiburg im Breisgau and Venice to Constantinople. The US failure to keep pace with the Europeans in terms of technology development would spur a significant postwar effort to be better prepared next time. Mitchell penned in his memoirs that he 'was sure that if the war lasted, air power would decide it' and detailed an aerial offensive plan echoing much of Douhet's ideas entailing a combination of incendiary fire and poison gas attacks aimed to destroy crops, forests, and livestock that 'would have caused untold sufferings and forced a German surrender' (Clodfelter, 1997: 87). For the remainder of his career, he vociferously promoted a progressivist version of military reform as an alternative to attrition warfare, one that would require an independent air force that could reach beyond the trenches to deprive an enemy the material capacity to wage war and to terrorize its citizens into overthrowing their own governments. Mitchell was considered as contentious in the US military as Douhet was in Italy; both were court-martialled for insubordination and for openly promoting their beliefs to the public, though unlike Douhet, Mitchell's court-martial was never overturned. He died in 1936 and would not live to experience any form of vindication when his plans would eventually come to fruition in the incendiary bombing attacks in subsequent wars over Japan, Korea, and Vietnam.

1.3 Imaginaries of a collective national self-defence (1920s–1930s)

Though the aviation component of the US Army employed the preponderance of its aircraft for reconnaissance missions in support of ground offensives during the First World War, airpower advocates – Mitchell, Gorrell and other military aviators, including Arnold – strove to demonstrate the airplane's decisive strategic ability to win wars independently from tactical engagements on the ground or at sea. A small group of adherents of Mitchell's ideas on strategic bombardment emerged during the 1920s as the core faculty of a new school, eventually designated the Air Corps Tactical School, that trained air component officers in air tactics and airpower theory. By the 1930s, the curriculum focused heavily on strategic bombardment theories. The most vocal core of instructors became known pejoratively as the 'Bomber Mafia' by their critics amid the Navy and Army, including other aviation specialties within the air component. They all would become influential general officers during the Second World War and beyond. In their efforts to carve out new roles and missions for the fledgling air forces, they strategized to intrude into offshore continental defence, which was traditionally the responsibility of the Navy. According to retired USAF colonel and security studies scholar Peter Faber, they deliberately 'sought to (1) define a threat, (2) repudiate the Navy's ability to answer the threat, and (3) offer a bomber-based solution' (Faber, 1997: 193). The point to stress here is that the threat that they conjured up, and the associated fear-mongering that ensued, had more to do with political manoeuvring and competition for influential power and resources than with any indication of a real threat to national security. These airpower advocates systematically redefined the concept of 'defensive' air operations to encompass initially the continental US, then gradually the hemisphere to protect the so-called sphere of influence outlined in the Monroe Doctrine[4] that has been a pillar of US foreign policy since 1823, and, finally, the world to target an enemy's 'vital' industrial centres no matter how far away they were.

Archived testimony statements that members of the 'Bomber Mafia' made before the Federal Aviation Commission in 1935 have documented some of the threats cited in support of their arguments for more comprehensive roles and responsibilities for airpower, characterized 'as nothing less than an anarchic, unregulated future' (Faber, 1997: 193). They employed language of an impending crisis of instability and disorder that could directly challenge US interests and that, in their minds, should be prevented with airpower. Major European powers – some potential future enemies – were outpacing the technological capabilities of the US and placing significant importance in developing air forces that they posited could potentially attack the vulnerable industrial capacity of the US, destroying railways, oil refineries, electric power systems and water supply systems that would quickly and efficiently destroy the people's euphemistic 'will to resist'. They contended that the US's primary centre of gravity was no longer its sea lanes of communication that had for so long secured its economic interests, but the industrial heartland of the nation now threatened by long-range strikes delivered via airpower.

[4] In his annual 'State of the Union Address' to Congress in 1823, President James Monroe outlined a policy designating the rest of the Americas as a US 'sphere of influence'. He asserted that the former Spanish colonies should remain free from further European influence and that any European intervention in the region would be viewed as a dangerous threat to US security. The policy was reanimated during the Cold War era of 'great power competition' with the Soviet Union. Some foreign policy pundits and government officials have asserted its relevance again in a contemporary competition with China.

They depicted a war-gaming scenario in which naval and land forces would be engaged, respectively, in Caribbean waters to defend the Panama Canal from a 'coalition of European and Asiatic powers' and in the western part of the US to defend a land invasion from the so-called 'Orient' (Faber, 1997: 194). According to Faber, they further asserted that this scenario would leave the industrial centre of the country open to undefended air attacks from the east via Canada that could only be defeated by an offensive projection of airpower. Their arguments in 1935 marked an early blurring of the spatial and temporal boundaries of US airpower. They projected a sociotechnical imaginary of a collective US national self-defence based upon a speculative war-gaming scenario of anticipated threats not founded on any degree of certainty, but on the modelling of possible alternative futures that could best be prevented in the present with airpower. These aspects of their strategy remain directly relevant to contemporary characterizations of threats to national security that will be analysed in the section on imminent threats. First, however, these early imaginaries of national security threats required more tangible manifestations so that they could be kinetically targeted by airpower.

1.4 Imaginaries of vital organs in the US precision bombing doctrine of the Second World War (1930s–1940s)

In preparation for the next war, the 'Bomber Mafia' developed a doctrine for high altitude precision daylight bombing (HAPDB), in which unescorted bomber aircraft would attack an enemy's perceived 'vital' economic industrial centres from *high altitudes*, with *precision*, in *daylight*. They preferred high altitude and daylight operations simply because technology did not yet enable safe and reliable beyond-visual operations. Anthropologists Stephen Collier and Andrew Lakoff have detailed through extensive archival research that these air warfare planners touted precision because it underpinned their theory that aerial destruction of a few essential nodes within what they described as 'national organic systems' of an enemy nation would lead to paralysis. Collier and Lakoff have explained that these airpower theorists analogized an enemy nation as a living organism that could be incapacitated by striking vital organs (Collier and Lakoff, 2021: 63). As one HAPDB planner asserted, precision bombing served as 'an instrument which could cause the collapse of this industrial fabric by depriving the web of certain essential elements – as few as three main systems such as transportation, electric power, and steel manufacture would suffice' (Collier and Lakoff, 2021: 64). The primary aim of precision targeting was therefore never intended to minimize the impact on civilians; it was designed with efficiency in mind to optimize the cost and effort to destroy as few nodes as necessary so that 'the whole of the economic machine ceases to function' (Collier and Lakoff, 2021: 64). Yet, targets of precision posed a problem because the 'Bomber Mafia' lacked knowledge about which essential nodes those might be. They primarily modelled plans to disrupt Germany's industrial infrastructure on more readily available data of US infrastructure, such as the 1920s Depression-era New Deal plans for New York City produced in support of social engineering and regional planning interventions. In the absence of any real knowledge about German urban planning and industrialization, they projected models of US planning onto Germany and assumed that the two were interchangeable.

HAPDB doctrine remained a key part of USAF operations until well into the US air war over Vietnam, when technology had advanced enough for the USAF to deem lower-altitude attacks less risky for its aircrews. It is worth questioning whether the now admitted failure of that doctrine had more to do with reactionary pressures against its destructiveness than with any misgivings in the logic behind its goals. The inaccuracies of HAPDB target selection were blamed on a dearth of intelligence sources, which would spur significant initiatives in ISR technology development over the coming decades, as well as concerted efforts to automate target recognition that will be further addressed in the section on identifying the enemy.

1.5 Imaginaries of cybernetic feedback loops and concentric rings (1970s onwards)

US air warfare theorizing since the war in Vietnam has promoted the same concept of rapid strategic paralysis underlying the war planning of these earlier airpower advocates albeit with different emphases. Contemporary theories have moved away from a focus on disruption of the enemy's economic capacity in hopes of affecting the psychological morale of the population's so-called 'will to resist' towards a focus on rapid psychological incapacitation of the enemy's ability to control and sustain its fighting forces (Fadok, 1997). When the airpower advocates of the 1930s described an enemy as a national living organism that could be incapacitated by targeting its vital economic organs, they drew upon emerging cybernetically oriented scientific theories that addressed cognition, behaviour and communication within a systems perspective. Airpower advocates since then have continued to embrace this systems thinking and incorporated their interpretations of contemporaneous complexity, chaos and network theories into their treatises. Colonel John Boyd, a fighter pilot who flew combat missions in 1953 during the war in Korea and commanded combat support units in the early 1970s in Vietnam, extensively synthesized these theories into his musings with titles such as 'Destruction and Creation' and

'The Essence of Winning and Losing'. He developed a four-step rational decision-making model, the Observe-Orient-Decide-Act (OODA) loop, as depicted in Figure 4, that has become a key component of current US warfighting doctrine.[5]

This theorizing assumes the centrality of information superiority in all operations. However, this superiority does not aim at attaining any genuine understanding of an adversary; it focuses on reaction times to target and disrupt an adversary's perceived decision-making OODA loop before that adversary can influence one's own. It dictates a mentality of pre-emption to avert risks to one's own OODA loop and, therefore, one's competitive advantage. It demands speed – the faster opponent supposedly wins. As international relations scholar Antoine Bousquet has indicated, such a strategy rests upon an ideology of endless conflict and an imperative for survival that advocates ambiguous and unpredictable action alongside perpetual change in order to keep the enemy disoriented and too paralysed to react (Bousquet, 2022: 183).

Colonel John Warden, another fighter pilot who flew in 1969 during the war in Vietnam and published a book on air campaign planning in 1989, also significantly influenced USAF airpower theorizing through his 'Five Rings' model (see Figure 5) aesthetically depicted as a bullseye of concentric 'vital' enemy centres of gravity with its leadership at the centre. The ultimate goal of this strategy is to sever the ability of the enemy's leadership to control its fighting forces, thereby inducing a physical *system* paralysis so that it can no longer effectively function. If that is not feasible, then the outer concentric rings should be targeted in order to induce 'unbearable psychological pressure on its leadership' to capitulate (Fadok, 1997: 373).

Warden led the air war planning for the First Iraq War (1991), and his adherents have continued to apply his model of systemic strategic paralysis in the initial airstrikes in Afghanistan in 2001 and the Second Iraq War (2003–2011), as well as drone strikes outside declared warzones against so-called terrorist networks.

The remainder of this chapter transitions to an analysis of themes within contemporary USAF narratives derived from these past imaginaries. The next section analyses the concepts of disorder, insecurity and crisis that are deemed disruptive and threatening to the USAF's sense of superiority, in turn driving an imperative to engage in an AWS-related arms race. It demonstrates how US airpower has been steadily utilized to expand the definition of national self-defence to encompass pre-emptive offensive attacks against purported imminent threats.

[5] Documentation of John Boyd's writings, briefings, talks and personal papers are available from: coljohnboyd.com.

Figure 4: John Boyd's OODA Loop, from 'The essence of winning and losing' (Boyd, 1995: 3)

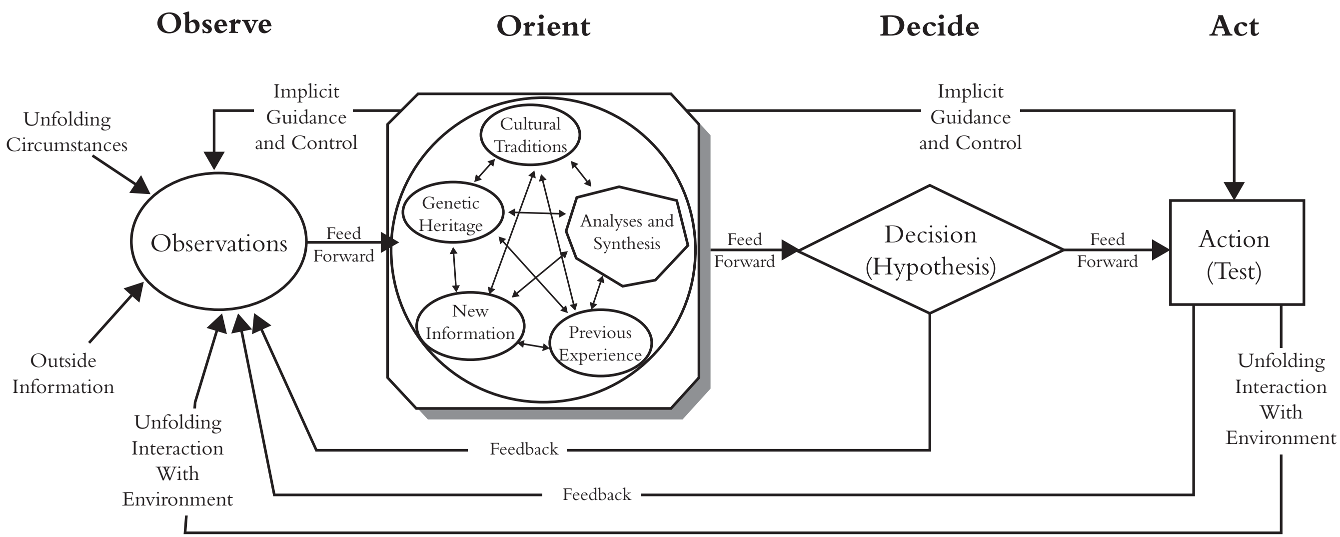
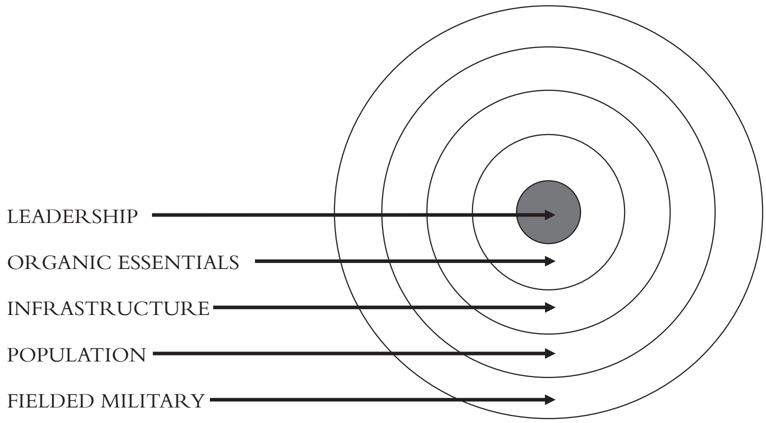


Figure 5: John Warden's five concentric rings



2. Imminent threat: omnipresent enemies and impending attacks against the homeland (rationalities from the 2000s onwards)

> When you think about it, everything and everybody has to be somewhere. (Retired USAF Lieutenant General James R. Clapper, as quoted in Ackerman, 2004)

> For the object, naturally, need not be there, it is enough that somewhere it *exist*: It is a possibility. Such an object is endowed with evil intentions and with all the attributes of a malefic power. (Fanon, 1986 [1952]: 120)

Arguments made by senior airpower leaders today to substantiate the necessity for significant investment in and employment of AWS technologies echo the language and motivations employed by their predecessors. While serving as Director of the DoD's National Geospatial-Intelligence Agency in 2004, retired USAF Lieutenant General James Clapper stated: 'Instead of one profound military threat, the United States and its allies face a global threat that can emerge almost anywhere' (Ackerman, 2004). More recently, Lieutenant General Timothy Haugh, a former commanding general of the USAF organization responsible for information warfare (IW), wrote in a military journal: 'As traditional organized power structures erode, disorder fills the void. We are moving from successive regional conflicts to a future characterized by continual global competition … Our adversaries have brought strategic competition to the nation's front door by engaging the [US] population in the information environment' (Haugh et al, 2020: 29). The previous USAF

Chief of Staff, General 'CQ' Brown, Jr.,[6] similarly declared in his 2020 guidance document *Accelerate Change or Lose*:

> Competitors, especially China, have made and continue aggressive efforts to negate long-enduring U.S. warfighting advantages and challenge the United States' interests and geopolitical position. While the Nation was focused on countering violent extremist organizations, our competitors focused on defeating us … Future warfare will not remain far from our shores; we must be prepared to address our competitors' attempts to hold the U.S. Homeland at risk with unconventional, conventional, and even nuclear force. (Brown, 2020: 3)

Such claims of disorder and insecurity are often taken for granted as objective statements of reality, but international relations scholar Didier Bigo has pointed out that these threats are socially constructed as problems by those responsible for providing for security. Their 'functionalist' or 'empiricist' vision of insecurity never includes the interactive roles they themselves play in creating the problems (Bigo, 2001: 92). Bigo has noted a conceptual triangle of identity, border and order at the centre of this political rationality of insecurity that rejects all difference and transforms it 'into a symptom of the undermining of a homogeneous societal identity as fantasized by the groups that declare its existence', one in which 'social and historical changes are perceived as a threat' and 'a structural phenomenon is transformed into an adversary' (Bigo, 2001: 100). The words of psychiatrist and political philosopher Frantz Fanon more than 50 years ago, in the context of his experiences in colonial France and the war for independence in Algeria, remain salient. He wrote of 'phobogenic objects' whose mere possibility of existence stimulates anxiety, fear and revulsion in those who cannot understand the diverse worlds of others, refuse to respect them, and attempt to exert a dominating universalizing cultural authority over them (Fanon, 1986 [1952]: 117).

2.1 Imaginaries of existential crisis and disruption to a normative world order

Just like the early airpower advocates' description of the threat environment a century ago, these 21st-century versions also contain elements of othering, spatializing and ordering. They reflect a racialized universalizing discourse

[6] General Charles Q. Brown, Jr. served as the Chief of Staff of the USAF from August 2020 to September 2023. In October 2023, he became the 21st Chairman of the Joint Chiefs of Staff, the most senior military officer in the US Armed Forces.

of global politics that orders 'threatening others' in relation to perceived positions within – and outside – a normative international system of states. General Brown depicts a hierarchical ordering that places the US at the top of this dominant spatiotemporality, China as an aggressive emerging competitor drawing the US into conflict, and the euphemistically labelled 'violent extremist organizations' (VEOs) as entities not worthy of inclusion as competitors. This description takes part in a larger civilizational discourse of friend-enemy and civilized-barbarian tropes (Walker, 2006: 68) with a long history that has included the labels of 'unwilling or unable', 'failed states', 'rogue regimes' and an 'axis of evil'.

General Brown's memo hinted at a possible sense of regret over a singular focus on so-called 'VEOs' over the past couple of decades that allegedly led to a loss of control over what was assumed to be exclusive access to superior technology, in turn implying that accelerated investment in *disruptive* technologies is now an existential imperative. Thus, he constructed a moment of crisis in which China's pursuit of the means to deny the USAF its ability to achieve air and information superiority requires significant political mobilization to change current USAF practices across technology acquisition, weapon systems development, and operations.

Anthropologist Janet Roitman (2014) has encouraged questioning what meaning is produced through the social construction of crisis narratives, in which systemic dichotomies and existing hierarchies are perpetuated. An understanding of the political work performed in such discourse helps to make visible their normative premises. Roitman's philosophically oriented analysis draws from sociologist Niklas Luhmann's work on systems theory in which he distinguished between first- and second-order observations. Roitman explains that claims to crisis are necessarily second-order observations because they produce meaning beyond first-order observations that something exists (beyond stating a first-order observation that an object is or is not a table, for example, a second order observation differentiates how that object is observed as a table in relation to a chair). Treating claims of crisis as first-order empirical observations and taking such assertions at presumptive face value results in diagnosis and critique that can only follow after the point in which such claims are made. In other words, an unchallenged judgement of crisis prompts questions about 'What went wrong?' only after that point of distinction – rather than asking why, how, what has changed or even 'What went right?' – since the very same systemic and structural conditions of the existence of an alleged crisis were considered productive and part of the normative way of doing things prior to its pronouncement (Roitman, 2014: 92–94, 99).

While the 'war on terror' has consumed USAF operations for more than two decades as a normative response to alleged 'terrorist networks', General Brown has now blamed VEOs for distracting the USAF's attention from its competitors and China for taking advantage of those circumstances in order to develop capabilities that can deny USAF air and information superiority. Such crisis narratives presuppose a failure – a negative integration – on the part of those blamed for the crisis and named as threats (in this case, 'VEOs' and China) and create blindspots that invoke some solutions (an AWS-related arms race) while foreclosing others. International relations scholars Jamie Johnson, Victoria Basham and Owen Thomas have likewise offered insights into the logic behind an appeal to crisis as '*a harm to the arrangement of privilege*' (Johnson et al, 2022: 616, emphasis in original). They argue that the politically oriented 'sense-making frames' and responses to the attack on 11 September 2001 represent 'a paradigmatic case of the disordering of order and the ordering of disorder' within international relations (2022: 609), in which the normatively constructed social order of where and to whom violent harm is expected to occur becomes disrupted.

2.2 Spatial and temporal imaginaries of an anticipatory national self-defence

The initial retaliatory air strikes in Afghanistan against al-Qaeda and the Taliban in response to the perceived existential disruption of the normative world order that '9/11' wrought have preceded what is now more than two decades of pre-emptive wars and 'over-the-horizon' drone strikes across Southwest Asia. Justifications for these offensive operations, framed within Article 51 of the UN Charter language of national self-defence and linked to the IHL principle of military necessity, resemble those made by their airpower predecessors against anticipated threats almost a century ago. Sociologist Lisa Stampnitzky has traced the first public use within US policy circles of the phrase 'pre-emptive action' to 2002 by then-President George W. Bush (2013: 109–110). While this reflects the first time that the term 'pre-emption' was publicly spoken by a US government official within the context of counterterrorism policies, a longer view on continuity within the logics of US airpower brings forth similar arguments made after the Second World War to validate the decision to drop atomic bombs over Japan in a pre-emptive 'first use' policy. In the words of then Secretary of War Henry L. Stimson, 'it was vital that [Germany] should not be the first to bring atomic weapons into the field of battle' and it would be 'a great new instrument for shortening the war and minimizing destruction' (Stimson, 1947: 102). General Arnold's injunction, cited at the opening of this chapter, that the US must never relinquish the means of preventing an attack against it remains equally instructive for interpreting the arguments of General Brown, his 22nd successor from August 2020 to September 2023, regarding the motivations for developing and employing AWS-related technologies.

The spatiotemporal notion of imminence that blurs the limits of what constitutes the proximity of a threat in both space and time can also be traced to early airpower advocacy. The underlying political logic has remained remarkably constant, though the technological means to distinguish threats and the legal means to justify their elimination have changed. The methodical redefinition of 'defensive' airpower operations to encompass any point in the world has delimited the physical aspect of US claims to act in national self-defence. The war-gaming scenario presented before the Federal Aviation Commission in 1935 posed an early manifestation of the anticipatory element in pre-emptive airpower planning that has also gradually delimited the temporal near-future aspect of those claims. The current operative definition of imminence publicly outlined by former Attorney General Eric Holder asserts – within the context of pursuing purported terrorist networks – that a defensive response is merited within a 'relative window of opportunity to act' in order to prevent *potential* attacks because their leaders are 'continually planning attacks against the [US]' (US Department of Justice, 2012).

Media theorist Mark Hansen has argued that in this definition of imminent threats:

> [W]hat is qualified – what actually comprises the threat – is not the likelihood of a particular event, of a single actual cause, coming to pass, but a far more complex and diffuse calculus of propensities concerning a myriad of factora [*sic*] that can only be known insofar as they can be qualified probabilistically. It is these micro-propensities, not the events they may go on to inform, that are the objects of probabilistic modeling. (Hansen, 2015: 109)

In an intelligence analysis textbook, US government defence contractors Patrick Biltgen and Stephen Ryan describe this process of probabilistic modelling as an automated human-machine teaming technique that extracts data from surveilled activity in order to make a potential threat more readily legible. They define what Hansen terms 'micro-propensities' as anomalous behaviours that differ slightly from a statistical model of what analysts expect to be typical or, rather, normative behaviours (Biltgen and Ryan, 2016: 221). As Claudia Aradau and Tobias Blanke (2022), scholars collaborating at the intersection of international relations, political theory, science and technology studies, and informatics, have shown – by following philosopher Achille Mbembe's notion of 'nanoracism' – this process of anomaly detection embeds racial inequality within the algorithmic reasoning that decomposes and recomposes data into figures of the enemy.

The following section focuses on this aspect of the political rationality that aims to exploit AWS-related technologies for an improved situational awareness. It explains the motivations for integrating AI-driven automated target and pattern recognition technology into anticipatory ISR operations focused on identifying potential threats. It highlights the relational uncertainty produced in anomaly detection that obfuscates the contingencies and indeterminacies involved in identifying allegedly suspicious and threatening activities.

3. Identifying the enemy with 'reasonable certainty'

> Our investigation now concludes that the strike was a tragic mistake. First, I will stress this was not a rushed strike. The strike cell deliberately followed and observed this vehicle and its occupants for eight hours while crosschecking what they were seeing with all available intelligence to develop a *reasonable certainty* of the imminent threat that this vehicle posed to our forces. (United States Department of Defense, 2021b, emphasis added)

> Distinguishing between friends and enemies is one thing; identifying the enemy with accuracy is quite another. (Mbembe, 2019: 49)

After three weeks of defending the 29 August 2021 drone strike in Kabul as morally 'righteous' (United States Department of Defense, 2021a) in 'eliminating an imminent ISIS-K threat' (US Central Command, 2021) against US forces, senior military leaders admitted that the intelligence was wrong (United States Department of Defense, 2021b). The drone strike killed Zemari Ahmadi – an employee of a US humanitarian aid organization – and nine members of his family.[7] The US military employs the ambiguous criteria of 'reasonable certainty' in its policy guidance governing targeting decisions related to the IHL principle of distinction that requires verification that an 'object of attack is a legitimate military target' (United States Department of Defense, 2009: GL-7). General McKenzie asserted it as a plausible defence for the incorrect conclusions drawn during eight hours of ISR operations. Even as evidence began to emerge that civilians had been killed in the airstrike, he described Ahmadi as an 'ISIS-K [terrorist] facilitator' (United States Department of Defense, 2021b), a vague classification that usefully serves to blur the distinction between a combatant and noncombatant. To reiterate Achille Mbembe's articulation of this condition, 'identifying the enemy with accuracy' is an elusive exercise when imaginaries invoke them everywhere and imbue them with the ability to threaten one's existence at

[7] The other victims include three of Ahmadi's children, Zamir, 20, Faisal, 16, and Farzad, 10; his cousin Naser, 30; three nephews, Arwin, 7, Benyamin, 6, and Hayat, 2; and two nieces, Somaya, 3, and Malika, 2.

any moment (Mbembe, 2019: 49). The alibi of 'reasonable certainty' leaves open juridical space for a lack of clear evidence and the possibility of doubt.

The pressures of public disapproval when such wrongdoings or 'scandals' become evident force the US military to issue ritual apologetic statements of temporary error to be corrected going forward, while still maintaining claims of legitimacy for its actions (Johnson et al, 2022: 619). In this instance, the USAF Inspector General presented three recommendations to address the 'tragic mistake': improved lateral communication between strike cells and their external supporting elements in order to enhance information sharing and situational awareness; a dedicated 'threat team' function to overcome confirmation bias – someone present to question preconceived notions that lead to misinterpretation of benign activity as suspicious behaviour; and a more thorough process to help better identify non-combatants present in potential strike areas (United States Department of Defense, 2021c). In this vein, the USAF promotes AI-driven automated pattern and target recognition technology as a means to reduce risk of harm to civilians due to its computational capacity to sift through significant amounts of data faster than humans with purportedly better accuracy to identify noncombatants.

3.1 Imaginaries of 'unknown unknowns': the fabrication of enemies

The USAF refers to its ten-year *Next Generation ISR Dominance Flight Plan* investment strategy framework as a culture change that will transform the platform-centric ISR operations of yesteryear into a problem-centric focus. The history of the platform-centric approach extends back to the inception of concerted efforts after the Second World War to develop the technological means to fill gaps in information – unknowns – that hampered the anticipated success of the 'Bomber Mafia's' HAPDB doctrine. The need to know what to target for an optimized *precision* effect drove significant investment in technical reconnaissance platforms to overfly enemy territory or otherwise access enemy information undetected that continues today. Those reconnaissance platforms have required a labour-intensive tasking process for an optimized allocation of relatively limited resources that focus on gathering precise amounts of information in service of narrow information requirements. The shift over the last 30-plus years since the First Iraq War to more high-volume approaches that require dedicated collection platforms, such as drones, for persistent surveillance and tracking of suspected threats active over an extended period of time have overburdened the tasking process. Biltgen and Ryan have explained that '[r]ather than seeking answers to predefined intelligence needs, collection attuned to [a problem-centric] methodology demands seeking data, in order to discover correlations and entity resolution' in activities deemed to be suspicious (Biltgen and Ryan, 2016: 133). In other words, this form of anticipatory intelligence moves away from searching for data to answer 'known unknowns' to searching for purported 'unknown unknowns' to be discovered in the vast amounts of captured data.

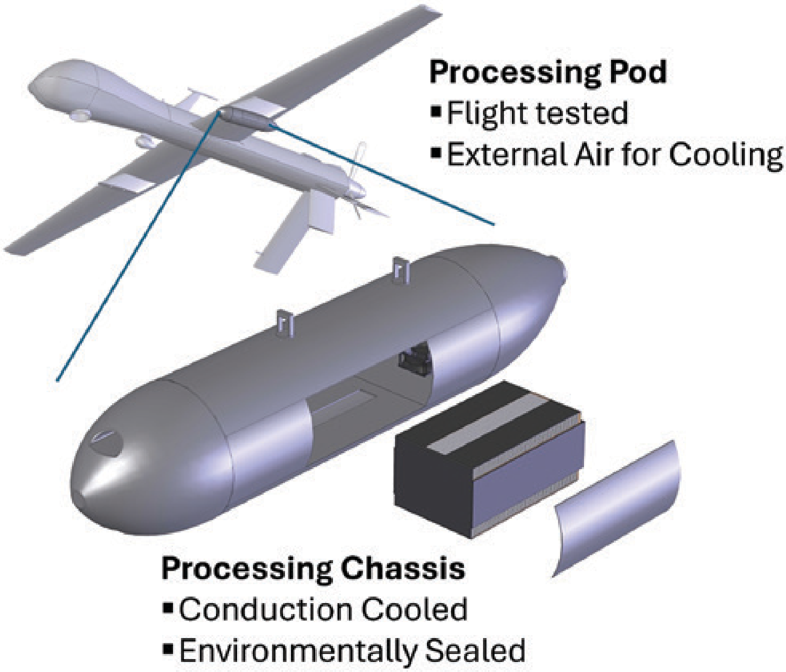
The transformation outlined in the USAF ISR dominance flight plan proposes a significant disruption in the ways in which intelligence-gathering platforms are tasked, in the ways in which the collected data are spatially and temporally correlated, and in the ways in which data are then fused, analysed and disseminated. It is within these practices that AWS-related pattern and target recognition technology is now being employed. The Agile Condor processing pod, designed to 'automatically detect, categorize, and track potential items of interest' (Trevithick, 2020), was originally flight-tested on a USAF drone in 2020 and has been adapted to additional air, land and sea-based platforms (SRC, Inc., n.d.). The pod has integrated neuromorphic computing hardware that provides teraflops of image and video processing power to enable pattern recognition along with an onboard data storage capability for edge computing, thereby eliminating the need to transfer the data to a location to be processed by humans (see Figure 6).

The Secretary of the USAF announced in 2021 that 'AI algorithms' had been employed for automated target recognition in a 'live operational kill chain' and that such technology had 'significantly reduced the manpower-intensive task of manually identifying individual targets, shortening the kill chain and accelerating the speed of decision making' (Miller, 2021). The desire is that humans will no longer spend hours staring at video surveillance footage since the controversial Project Maven[8] has created algorithms that classify that video data, curate and then label their contents to create suitable datasets for machines to process through an abductive logic and to alert humans when certain patterns emerge.

Political geography scholar Louise Amoore and digital humanities scholar Rita Raley have demonstrated the political work involved in the abductive logic of AI, in which machine learning algorithms proceed backwards from a solution – the elimination of a threat, for example – to define the problem,

[8] The existence of the DoD's Project Maven to integrate AI technologies into analysis of drone video footage – and Google's initial collaboration on it – was first publicly reported in 2018. Google employees signed an open letter urging company leadership to cease collaborating with the DoD on weaponized AI technology. Palentir Technologies assumed industry lead shortly thereafter. The project led to the establishment of the DoD's Joint Artificial Intelligence Center (JAIC) that served as the focal point between 2018 and 2022 for coordinated integration of AI technologies across the US military. The JAIC was dissolved in June 2022 and its responsibilities became subsumed into the roles of DoD's Chief Digital and Artificial Intelligence Officer. Oversight for Project Maven also transferred in 2022 from the Pentagon's Office of Undersecretary of Defense for Intelligence and Security to the National Geospatial-Intelligence Agency, signalling an operationalization of its control.

Figure 6: Illustration of Agile Condor processing chassis and pod mounted under the wing of a drone platform[9]



the potential threat (Amoore and Raley, 2017; Amoore, 2023). The Agile Condor computing processors can generate potential threats in this manner within a matter of seconds via recognition and correlation of patterns among anomalous activities across significant amounts of disparate data. Such computational work actively forecloses possibilities that the observed entities are not a threat. In addition to automated target recognition technologies like Agile Condor, query capabilities based on predictive language foundation models and multimodal generative AI technologies will sort through all available data most relevant to analysts' requests for information, and they will 'learn' over time to push tailored notices to them (Seffers, 2018). All of this is intended to free humans to assess the decomposed and recomposed data, and make decisions in increasingly tighter time loops, faster and faster inside the enemy's perceived decisional OODA loop. USAF intelligence personnel now refer to themselves as 'sense makers' who will converge on problems as teams with a swarm mindset (Borukhovich and Morton, 2020).

[9] The illustration of the Agile Condor processing pod is available on the SRC Inc website: srcinc.com/products/intel-collection-and-analysis/agile-condor-highperformance-embeded-computing.html.

3.2 Imaginaries of an elusive situational awareness

The USAF asserts that integration of AWS-related technologies into its ISR operations will deliver not only improved situational awareness, but also a deeper 'situational understanding' (Kimmons and Gilmer, 2019: 5) that is supposed to eliminate the so-called 'tragic mistakes' that have led to the killings of the Ahmadi family and untold others. However, Amoore and Raley stress that human analysts cannot fully discern how the deep learning neural networks of AI technologies identify activities to surface for their attention. Therefore, analysts cannot critically assess the associative meaning embedded in prompts presented by these algorithms. Their incorporation of these abductively generated inputs from AWS-related technologies into their lethal decision-making processes – in terms reminiscent of Roitman's crisis-critique analysis outlined earlier – amounts to accepting at presumptive face value the second-order observations of AWS technologies. Any questioning – critique – of that automated input begins from a presumption that the anomalous activity is a potential threat, already suppressing any possibility that it is not.

These human-machine practices do not translate data into any genuine knowledge about particular so-called threats or even into an accurate portrayal of situational awareness, because the data collected and processed by this sociotechnical assemblage are inherently partial, biased and doubtful. Such knowledge is based on a 'conjectural style of reasoning' (Aradau and van Munster, 2011: 31–51) that extrapolates and reconstructs collected data from the reality of their social contexts, thus reducing suspect persons to mere data points to be analysed for discernible patterns and connections of normatively determined 'abnormal' behaviour among the smallest of details that are deemed to be threatening. Furthermore, AWS-related technologies 'learn' to predict temporal relations between the input data and a set of forecasted values based on algorithmic models that contain significant magnitudes of hidden layers, weighted probabilities and potential connections with little to no account for uncertainty in the data (Amoore, 2020; Aradau and Blanke, 2022; Amaro, 2023). The point to be made here is that automated target and pattern recognition technologies like Agile Condor do not work to eliminate uncertainty; they serve to mirror and amplify in time and space the human judgements that employ them. The USAF Inspector General's recommendations for corrective actions amid public evidence of faulty intelligence in the Kabul drone strike promise improved accuracy in distinguishing between combatants and noncombatants, but the practices and technologies at work within USAF ISR operations instead generate what could be characterized as an overall situational unawareness.

Conclusion: meaningful human control

This chapter has established more than a century-long continuity of sociotechnical imaginaries within US military airpower advocacy and employment that seek to induce rapid strategic paralysis of enemies. The notion of disruption has constituted a recurring theme within this air-mindedness, employing new technologies with aims to generate quick psychological victories over enemies and thwart any plans to employ those technologies against the US in a similar manner. These imaginaries have been braced by a political rationality that has assumed a privileged superiority in the knowledge, means and moral authority to exploit the destructive potential of technology. Lethal AWS-related technologies incorporating AI that can compute exponentially greater amounts of data at significantly faster speeds than humans hold a captivating *allure* for strategies that seek to defend this racialized and moralizing worldview and to impose their version of universalism on others. These technologies also offer *a lure*, a means to capture data that can be decomposed and recomposed in service of those coercive aims.

Though USAF ISR operations claim to get inside the heads of their opponents, they instead create sociotechnical imaginaries around an offensive, pre-emptive violence that negates the need to do so. This violence is fuelled by insecurities in the face of contestation of and challenges to the US' ability to influence the rules of the normative international order founded in systemic geopolitical and economic inequalities the US and its allies have worked to maintain to their advantage. This insecurity generates a proliferation of imaginaries of threats to a collective national security. Moreover, they are manifestations of global politics in which the actions of the US are thoroughly entangled and cannot credibly be depoliticized. US claims to a moral superiority that guides a benevolent intent in its actions compared to those of others, of its enemies, reflect an historical amnesia and neglect the deeply political presuppositions inherent in such claims of exceptionalism.

This chapter has problematized human judgement within the USAF, the 'meaningful human control' championed in regulatory debates on lethal AWS technologies to serve as a safeguard against 'killer robots' that could carry out lethal airstrikes without human oversight. It has endeavoured to demonstrate that all the human-machine interactions leading to a lethal decision are *meaningful* in the sense that they are imbued with a particular political rationality that steers their activities from the outset, irrespective of the degree of machine autonomy present at the lethal decision point. It has highlighted circumstances surrounding US airpower advocacy in which military objectives of efficiency and combat effectiveness have shaped interpretations of the key principles of proportionality, necessity and distinction within IHL that are intended to protect noncombatants during the conduct of war. The concept of precision targeting has likewise served to optimize military cost and effort in striking enemy targets, not to minimize the impact to noncombatants. This has enabled interpretations of IHL that seek to legitimize USAF strategic interests above those they harm and then excuse those harms as so-called 'unintended consequences' and 'tragic mistakes'. A more productive way to contest the security practices of USAF airpower and to regulate the use of AWS-related technologies effectively could begin with centring within IHL the experiences of the victims affected by them. Such an approach could work to displace the primacy of national security threat imaginaries to their very real and destructive consequences. It could de-emphasize a focus on permissibility within the laws of war and demand genuine democratic accountability for infliction of harm upon others. That, for me, seems a much better notion of 'meaningful human control'.

Acknowledgements

This chapter builds upon the blog article, 'Shortening the kill chain with artificial intelligence', *Autonorms*, 28 November 2021, available from: autonorms.eu/shortening-the-kill-chainwith-artificial-intelligence.

I would like to thank Ingvild Bode in the Center for War Studies at the University of Southern Denmark and Nadia Fadil and Alexandra Oancă in the Department of Social and Cultural Anthropology at the Catholic University of Leuven (KU Leuven) for their dialogue and generous support. I thank the editors of this volume for their professional feedback and guidance throughout the publication process.

References

Ackerman, R.K. (2004) 'Digitization brings quantum growth in geospatial products', *Signal*, 1 August. Available from: afcea.org/signal-media/digitization-brings-quantum-growth-geospatial-products [Accessed 24 January 2023].

Air & Space Forces Magazine (2021) 'VIDEO: Kendall on the State of the Forces at AFA's Air, Space & Cyber '21', 23 September. Available from: airandspaceforces.com/video-kendall-on-the-state-of-the-forces-at-afas-air-space-cyber-21/ [Accessed 26 May 2023].

Amaro, R. (2023) *The Black Technical Object: On Machine Learning and the Aspiration of Black Being*, Berlin: Sternberg Press.

Amoore, L. (2020) *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*, Durham, NC: Duke University Press.

Amoore, L. (2023) 'Machine learning political orders', *Review of International Studies*, 49(1): 20–36. https://doi.org/10.1017/S0260210522000031

Amoore, L. and Raley, R. (2017) 'Securing with algorithms: knowledge, decision, sovereignty', *Security Dialogue*, 48(1): 3–10. https://doi.org/10.1177/0967010616680753

Aradau, C. and Blanke, T. (2022) *Algorithmic Reason: The New Government of Self and Other*, Oxford: Oxford University Press.

Aradau, C. and van Munster, R. (2011) *Politics of Catastrophe: Genealogies of the Unknown*, New York: Routledge.

Arnold, H.H. (1945) *Third Report of the Commanding General of the Army Air Forces to the Secretary of War*, Baltimore: Schneidereith & Sons.

Bigo, D. (2001) 'The Möbius ribbon of internal and external security(ies)', in M. Albert, D. Jacobson, and Y. Lapid (eds) *Identities, Borders, Orders: Rethinking International Relations Theory*, Minneapolis: University of Minnesota Press, pp 91–116.

Biltgen, P. and Ryan, S. (2016) *Activity-Based Intelligence: Principles and Applications*, Boston: Artech House.

Borukhovich, K. and Morton, T. (2020) 'DCGS next generation: accelerating change to deliver decision advantage', *Over the Horizon*, 26 September. Available from: othjournal.com/2020/09/26/dcgs-next-generation-accelerating-change-to-deliver-decision-advantage/ [Accessed 10 May 2023].

Bousquet, A. (2022) *The Scientific Way of Warfare: Order and Chaos on the Battlefields of Modernity*, 2nd edn, Oxford: Oxford University Press.

Boyd, J.R. (1995) 'The essence of winning and losing', *John Boyd Homepage*. Available from: www.colonelboyd.com/boydswork [Accessed 3 November 2023].

Brown Jr., C.Q. (2020) *Accelerate Change or Lose*, Washington, DC: United States Air Force.

Clodfelter, M.A. (1997) 'Molding airpower convictions: development and legacy of William Mitchell's strategic thought', in P.S. Meilinger (ed) *The Paths of Heaven: The Evolution of Airpower Theory*, Maxwell Air Force Base, Montgomery, AL: Air University Press, pp 79–114.

Collier, S.J. and Lakoff, A. (2021) *The Government of Emergency: Vital Systems, Expertise and the Politics of Security*, Princeton: Princeton University Press.

Driant, E. (1911) *Au-dessus du Continent Noir*, Paris: Ernest Flammarion.

Faber P.R. (1997) 'Interwar US Army aviation and the Air Corps tactical school: incubators of American airpower', in P.S. Meilinger (ed) *The Paths of Heaven: The Evolution of Airpower Theory*, Maxwell Air Force Base, Montgomery, AL: Air University Press, pp 183–238.

Fadok D.S. (1997) 'John Boyd and John Warden: airpower's quest for strategic paralysis', in P.S. Meilinger (ed) *The Paths of Heaven: The Evolution of Airpower Theory*, Maxwell Air Force Base, Montgomery, AL: Air University Press, pp 357–398. http://dx.doi.org/10.21236/ada328849

Fanon, F. (1952/1986) *Black Skin, White Masks*, London: Pluto Press.

Gat, A. (1997) 'Futurism, proto-fascist Italian culture and the sources of Douhetism', *War & Society*, 15(1): 31–51. https://doi.org/10.1179/072924797791201021

Hansen, M.B.N. (2015) 'Our predictive condition; or, prediction in the wild', in R. Gruisn (ed) *The Nonhuman Turn*, Minneapolis: University of Minnesota Press, pp 101–138.
Haugh, T.D., Hall, N.J. and Fan, E.H. (2020) '16th Air Force and convergence for the information war', *Cyber Defense Review*, 5(2): 29–43.
Johnson, J.M., Basham, V.M. and Thomas, O.D. (2022) 'Ordering disorder: the making of world politics', *Review of International Studies*, 48(4), 607–625. https://doi.org/10.1017/S0260210522000183
Kimmons, J. and Gilmer, G. (2019) *Maintaining Advantage: Remaking PED for Today's Intelligence Needs*, McClean, VA: Booz Allen Hamilton Inc.
Maksel, R. (2011) 'The world's first warplane', *Air & Space Magazine*, 21 October. Available from: smithsonianmag.com/air-space-magazine/the-worlds-first-warplane-115175678/ [Accessed 29 May 2023].
Mbembe, A. (2019) *Necropolitics*, Durham, NC: Duke University Press.
Meilinger, P.S. (1997a) 'Introduction', in P.S. Meilinger (ed) *The Paths of Heaven: The Evolution of Airpower Theory*, Maxwell Air Force Base, Montgomery, AL: Air University Press, pp xi–xxx.
Meilinger, P.S. (1997b) 'Giulio Douhet and the origins of airpower theory', in P.S. Meilinger (ed) *The Paths of Heaven: The Evolution of Airpower Theory*, Maxwell Air Force Base, Montgomery, AL: Air University Press, pp 1–40.
Miller, A. (2021) 'AI algorithms deployed in kill chain target recognition', *Air & Space Forces Magazine*, 21 September. Available from: airandspaceforces.com/ai-algorithms-deployed-in-kill-chain-target-recognition/ [Accessed 29 May 2023].
Mitchell Institute for Aerospace Studies (2018) 'The Mitchell hour: The ISR flight plan roundtable', 2 August. Available from: mitchellaerospacepower.org/event/the-isr-flight-plan/ [Accessed 11 November 2023].
Patterson, I. (2011) '100 years of air strikes', *London Review of Books*, 21 March. Available from: lrb.co.uk/blog/2011/march/100-years-of-air-strikes [Accessed 15 May 2023].
Roitman J. (2014) *Anti-crisis*, Durham, NC: Duke University Press.
Salvaneschi, N. (1917) *Let Us Kill the War: Let Us Aim at the Heart of the Enemy!*, Milan: Art Edition of Bianco e Nero.
Seffers, G.I. (2018) 'In vast oceans of data, AI extracts the pearls', *SIGNAL*, 1 March. Available from: afcea.org/content/vast-oceans-data-ai-extracts-pearls [Accessed 29 May 2023].
SRC, Inc. (n.d.) 'Intel collection & analysis systems: Agile Condor® high-performance embedded computing'. Available from: srcinc.com/products/intel-collection-and-analysis/agile-condor-high-performance-embeded-computing.html [Accessed 29 May 2023].
Stampnitzky, L. (2013) *Disciplining Terror: How Experts Invented 'Terrorism'*, Cambridge: Cambridge University Press.

Stimson, H.L. (1947) 'The decision to use the atomic bomb', *Harper's Magazine*, February, pp 102–107.

Trevithick, J. (2020) 'MQ-9 Reaper flies with AI pod that sifts through huge sums of data to pick out targets', *The War Zone*, 4 September. Available from: thedrive.com/the-war-zone/36205/reaper-drone-flies-with-podded-ai-that-sifts-through-huge-sums-of-data-to-pick-out-targets [Accessed 29 May 2023].

Turse, N. (2021) 'Drone strike victims in Yemen are desperate for accountability from the U.S.', *Vice*, 15 April. Available from: vice.com/en/article/n7bj8b/drone-strike-victims-in-yemen-are-desperate-for-accountability-from-the-us [Accessed 4 November 2023].

United States Central Command Public Affairs (2021) 'U.S. Central Command statement on defensive strike in Kabul', 29 August. Available from: centcom.mil/MEDIA/Transcripts/Article/2781320/general-kenneth-f-mckenzie-jr-commander-of-us-central-command-and-pentagon-pres/ [Accessed 13 December 2023].

United States Department of Defense (2009) 'Chairman of the Joint Chiefs of Staff Instruction 3160.01, No-Strike and the Collateral Damage Estimation Methodology', *ACLU Drone FOIA – Department of Defense Documents* 13 February. Available from: aclu.org/documents/no-strike-and-collateral-damage-estimation-methodology [Accessed 29 May 2023].

United States Department of Defense (2021a) 'Transcript: Secretary of Defense Austin and Chairman of the Joint Chiefs of Staff Gen. Milley press briefing on the end of the U.S. war in Afghanistan', 1 September. Available from: defense.gov/News/Transcripts/Transcript/Article/2762169/secretary-of-defense-austin-and-chairman-of-the-joint-chiefs-of-staff-gen-mille/ [Accessed 13 December 2023].

United States Department of Defense (2021b) 'Transcript: General Kenneth F. McKenzie Jr. Commander of U.S. Central Command and Pentagon Press Secretary John F. Kirby hold a press briefing', 17 September. Available from: defense.gov/News/Transcripts/Transcript/Article/2780738/general-kenneth-f-mckenzie-jr-commander-of-us-central-command-and-pentagon-pres/ [Accessed 13 December 2023].

United States Department of Defense (2021c) 'Transcript: Pentagon Press Secretary John F. Kirby and Air Force Lt. Gen. Sami D. Said hold a press briefing', 3 November. Available from: defense.gov/News/Transcripts/Transcript/Article/2832634/pentagon-press-secretary-john-f-kirby-and-air-force-lt-gen-sami-d-said-hold-a-p/ [Accessed 13 December 2023].

United States Department of Justice (2012) 'Attorney General Eric Holder speaks at Northwestern University School of Law', *Justice News*, 5 March. Available from: justice.gov/opa/speech/attorney-general-eric-holder-speaks-northwestern-university-school-law [Accessed 26 April 2023].

Van Dijk, B. (2022) *Preparing for War: The Making of the Geneva Conventions*, Oxford: Oxford University Press.
Walker, R.B.J. (2006) 'The double outside of the modern international', *Ephemera: Theory and Politics in Organization*, 6(1): 56–69.
Warden, J.A. (1989) *The Air Campaign: Planning for Combat*, Washington DC: Brassey's.
Wohl, R. (1994) *A Passion for Wings: Aviation and the Western Imagination, 1908–1918*, New Haven: Yale University Press.

From *Maschinenmensch* to Robot Bubs: Female-Presenting Autonomous Weapons Systems in Live-Action Films 1927–2023

Rebecca L. Jones

Introduction

Few people associate femininity with autonomous weapon systems (AWS), yet live-action science-fiction (SF) films have depicted lethal female-presenting creations since the early days of cinema. Christof Heyns (2016: 4) defines AWS as 'robotic weapons that, once activated, can select and engage targets without further human intervention'. These AWS follow programmed directives, exhibiting autonomy only within set parameters. He distinguishes these from systems that lack 'meaningful human control over force release', which he considers 'fully autonomous' (2016: 6). Thus, with real-world violence commonly enacted by men, the idea of women being lethal is rarer and might be deemed less relevant, especially to discussions of AWS. Yet, female-presenting AWS (abbreviated here as F-PAWS) in live-action films infiltrate where their male counterparts more often blast their way in, creating the SF trend that if you want to infiltrate and access a target quietly, then an F-PAWS is used. Accordingly, F-PAWS operate within programmed limitations, typically as killer machines directed by men to use their feminine appearances to infiltrate society, seduce targets and then harm or kill them. These representations often present their male creator or target's idea of appealing femininity, a reductionist performance that speaks to a discordant combination of martial violence and femininity. Using Heyns' definition of AWS and full autonomy and applying it to the depiction of F-PAWS, this chapter analyses the representational types established by F-PAWS in live-action feature-length films to argue that these AWS' female presentations are a lethal, weaponized femininity revealing contemporary understandings of appealing femininity, women's place in conflict, and the fears and potentiality of human-passing AWS and autonomous military artificial intelligence (AI).

At the time of writing, there are no offscreen human-passing or human-appearing AWS, yet they have appeared in SF films since 1927, with some characters serving their controllers, while others rebel. Notably, control over AWS is central to current debates around their military employment. There are already many instances of AWS in use,[1] though none is fully autonomous by Hayns' standards. Significantly, the US Department of Defense (DoD) Directive 3000.09 (2017: 2) requires that AWS 'be designed to allow commanders and operators to exercise appropriate levels of human judgement over the use of force'. Thus, all DoD AWS must be controllable by humans. However, liability and responsibility for the actions of these creations are still up for debate and of great concern.[2] Alternatively, live-action films' F-PAWS are predominantly fully autonomous once sent on their missions, exemplifying Heyns' (2016: 6) 'no meaningful human control over [its] force release' distinction. However, there are cinematic instances controlled remotely through radio signals or command words, presenting a 'human on the loop' scenario, meaning humans can give commands and intervene with the robot's otherwise autonomous actions (Caton, 2015: 2). More recent films have F-PAWS who[3] are sentient agents surpassing, rejecting or never under the control of others, presenting 'human out of the loop' AWS, where the humans have no control or ability to intervene with the robot's actions (Caton, 2015: 2).

There exists no comprehensive study of F-PAWS depictions to date, ignoring the significance of these depictions within SF and the context of AWS. Thus, this analysis reveals how entrenched the idea of nonthreatening femininity[4] is through the repeated depiction of F-PAWS as primarily infiltrators, with the films' roboticists, and occasionally the F-PAWS themselves, using female presentations to conceal their violent purpose. Judith Butler's (2006) theory of gender performativity asserts that gender is learned and performed. Thus, AI, both onscreen and offscreen, is perceived

1 See Caton (2015) for a list of currently operational AWS.

2 See International Committee of the Red Cross reports (2014, 2016 and statement 2021) and Bhuta et al (2016).

3 I will be using 'who' to refer to the cinematic instances who present as sentient since they are depicted as persons rather than just objects (who/she) and 'it' to refer to the instances that never display any such development or capability (that/it).

4 Mainly concerning the US as most of the films with AWS are US productions, so the overall trends examined reflect US sentiments.

as feminine or masculine due to attributes like voice, appearance, function, or behaviour, and is referred to with gendered pronouns and given gendered names. Creators do this to foster positive and nonthreatening interactions between creations and users. Thus, by having F-PAWS perform femininity as submissive, nonthreatening and sexually appealing to heterosexual men, these characters speak to their respective cultures' understanding of femininity as such and what it sees as femininity's function within combat.

Ying Liang (2015: 2037) contends that the conventional female form was considered inherently frail and vulnerable, and this sentiment is reinforced by narratives favouring male action heroes, soldiers and other violent enactors. The films examined herein reflect these associations by repeatedly connecting femininity with emotionality and weakness, while also capitalizing on the expectation that beautiful women are physically unthreatening. F-PAWS's gender performance combined with these associations facilitate their onscreen infiltration and lethal seductions acting as cautionary tales for future offscreen F-PAWS functionalities and possible outcomes. By examining these representational patterns with the feminist understanding of gender as a performance, we gain an insight into how cultural perceptions and expectations surrounding what is appealing about femininity underpins F-PAWS' effectiveness and potentiality for violence. Additionally, these depictions speak to the gendered place of women within conflicts, showing often impossible lethal capability but also the viability of violent women.

This onscreen sexualization of F-PAWS also occurs in offscreen robots and AI where gender-presenting creations (male or female voice, name or appearances) are referred to with gendered pronouns. The study 'The public's perception of humanlike robots' by Megan Strait, Cynthia Aguillon, Virginia Contreras and Noemi Garcia (2017: 1423) found that 'the frequency at which [their subjects] sexualized the female-gendered robots eclipsed that of all other concepts examined'. Additionally, other studies found that, on average, men prefer interacting with female-presenting AI and robots to male-presenting ones.[5] These findings help explain how and why roboticists who wish their creations to seem more human and approachable make their AI and robots present as female when they are intended to interact with the wider public. This appeal to gender performativity reveals the disconnect between gender and organic form, as these creations need only a feminine voice or appearance for humans to gender them as female. Thus, this sexualization of female-presenting artificial creations both onscreen and offscreen speaks to how SF shapes and is shaped by technological realities, the performative nature of gender, and the perception of femininity as trustworthy and nonthreatening.

[5] Siegel, Breazeal and Norton (2009: 2566); Wosk (2015: 6).

There is a 96-year gap between the first instance of an F-PAWS and the last as of 2023, containing 42 live-action films[6] which respond to the increasing reality of AWS offscreen (none of which is currently capable of passing as human), the rise in female-presenting AI assistants (products of military funding)[7] and the changes in social expectations for artificial female-presenting creations (from sexual objects to lethal threats who no longer use erotic allure to achieve their goals). These fictional characters speak to how a human appearance does not guarantee personhood, but does show how we expect human behaviours from these creations and accept a human appearance and behaviour as signifiers of sentient identity. I argue that the representational and thematic evolution from the first instance of an F-PAWS in live-action films to the last as of 2023 signifies a transition from controlled seductresses to uncontrolled agents. This evolution is not linear, changing as new trends emerge while often returning and retaining established themes. These trends create character types which speak to ideas around femininity, violence, and current and future use of AWS showing the lethal potentiality but also the danger of creating things that can exceed our control.

The first section addresses this chapter's methodology and scope with subsequent sections discussing three types based on a taxonomy of representational trends. The next section addresses the *Lethal Seductress* type, gynoids (female-presenting robots)[8] and AI designed to sexually allure men, enhancing their effectiveness at infiltrating and seducing targets. Characters of this type initially present controlled, weaponized femininity, but over time sentient depictions increase, resulting in more rejections of their functionality and assertions of autonomous agency. The third section presents the second type, *Threatening Machines*: characters that eschew seduction as a means of infiltration; instead, their femininity exists only for aesthetic appeal. Instances of this type begin operating within their designed parameters, with later instances developing agency as their origins shift from military to extraterrestrial. Finally, the final section addresses the *Uncontrolled Agents* type: F-PAWS that are uncontrolled, non-eroticized, nonseductress, fully autonomous, sentient characters. Each section chronologically traces the evolution of its type and how their instances explore themes of weaponized,

6 That I found within my wider project's survey of 336 live-action films with robots and AI in them. See https://bookknight101.wordpress.com/2019/05/29/121-years-of-robots-and-ai/ for a full list of films viewed.

7 The first is Siri (2011, Apple Inc.), an AI voice assistant originally developed by the Defense Advanced Research Projects Agency (DARPA) (WELT, 2012).

8 These characters are human-appearing or human-passing robots with physical bodies designed to emulate *female* attributes. I use this term to signify characters with physical forms versus those who exist only within the digital realm, for whom I use the term 'AI'.

controlled femininity and uncontrollable technologies. The examined 42 films explore potential futures and outcomes for AWS, technologies and social conceptions of femininity. Though fictional, these narratives caution us about the consequences of losing control over AWS, the dangers of underestimating women, and the ways in which social norms infiltrate and impact our technological and martial landscapes.

1. Methodology and scope: understanding onscreen F-PAWS

This chapter seeks to present why inorganic AWS are created to present as female, how that femininity is weaponized, what this says about our contemporary understanding of femininity, the problematic associations these depictions form and respond to, and broader debates about AWS. Focusing on representations of F-PAWS, I explore femininity's potential weaponization, a trend not replicated within male-presenting AWS, because their masculinity is *not* employed as a strategic tool. In my broader analysis of 350 feature-length live-action films with robot or AI depictions (from 29 countries,[9] with the US being the main producer), 135 of the films feature *female*-presenting AI or robots. Eighty of those films depict Galatea-like[10] creations functioning as girlfriends, sexbots and seductresses versus the 43 featuring F-PAWS characters from the first in 1927 to the last as of 2023, 26 of which are included in the 80, highlighting the limited prevalence of femininity functioning as a violent threat and the greater emphasis on presenting idealized, controlled femininity. This underscores the necessity of delving into representations of F-PAWS and the insights these depictions provide in terms of contemporary and future AI and AWS technologies, and their possible implementation of femininity.

Most of the films within this chapter's scope originate from the US and the UK, consequently skewing this chapter's analysis towards those countries' SF representational trends.[11] The scope of this chapter is confined to live-action productions, excluding animated works, since onscreen depictions of this

[9] Australia, Belgium, Bulgaria, Canada, China, Columbia, East Germany, France, Germany, Hungary, Hong Kong, India, Italy, Japan, Mexico, the Netherlands, New Zealand, Puerto Rico, Poland, Romania, Serbia, South Africa, South Korea, Spain, Switzerland, the Union of Soviet Socialist Republics, the United Arab Emirates, the UK and the US.

[10] Referring to the Pygmalion myth, where a sculptor created a statue of everything he thought of as ideal in women, fell in love with it (taking it to his bed at night), and was rewarded with its animation by Aphrodite, who saw his love for his creation and so made it a human woman, whom he married.

[11] The four other countries which produced live-action films with F-PAWS are: Australia, Germany, Hong Kong and South Korea.

nature originated in live-action films;[12] as such, my longitudinal investigation commences with this format. This chapter uses categories developed during my doctoral research (Jones, 2022) where I created a taxonomy of overarching and overlapping categories that encapsulated the observed representational trends. This taxonomy changed and was revised by the end of 2023, ensuring that it accommodated the examined depictions and facilitated discussions of each trend's evolution across multiple, distinct films. From this, I identified the Lethality archetype and its three types: Lethal Seductress, Threatening Machines, and Uncontrolled Agents. I examined only machine instances because they lack organic attributes like sex organs and gender unless they are intentionally constructed and programmed with them. Thus, these characters reflect their in-film creators' or manufacturers' understanding of gender, their film's contemporary social perception of gender and AI technologies, and SF's exploration of femininity's application within AI and robotic creations. It is from this locus that I examine their contemporary contextual standards, expectations and perceptions, and extrapolate the F-PAWS fabular themes by applying Butler's theory of gender performativity to address each character's gender presentation as a performance that is a distillation of their creator's, film's and SF's contemporary understanding of femininity and debates around AI and AWS technologies.

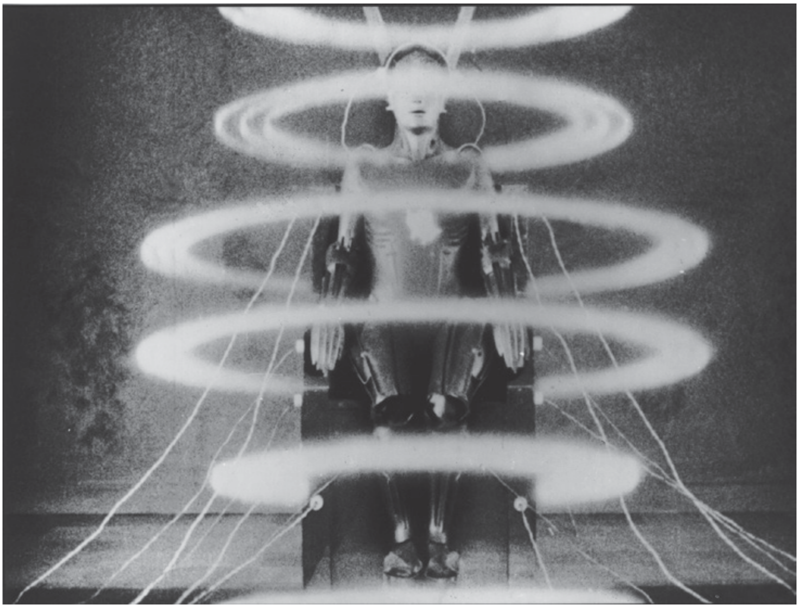
2. Lethal Seductresses: weaponized femininity

The first type, Lethal Seductresses, encompasses gynoids and AIs designed to be sexually alluring to enhance their ability to infiltrate or seduce their targets. A notable trend within this type is their high rate of destruction, primarily because, as Sue Short (2011: 98) notes, 'female sexuality and independence … is portrayed as terrifying and … is accordingly restrained' through their destruction. These artificial entities are assassins, killers and deceivers, often functioning as antagonists necessitating their elimination. Characters of this type are frequently devoid of sentience or self-awareness. Yet the characters that attain sentience often reject their primary objective, opting for romantic attachments instead.[13] In either case, they present AWS that defy their target's expectations through their feminine appearance, evincing the effectiveness of F-PAWS as infiltrators and the dangers of programming machines to simulate or have emotions.

[12] The first recorded instance of a mechanical, artificial creation onscreen was in 1897's *Gugusse et l'Automate* by Georges Méliès; the first surviving depiction of an AI or robot on-screen is *The Inventor's Secret* (1911), a nickelodeon directed by Mack Sennett.

[13] No 7 (*Some Girls Do*), Andrea⋆ (*Angel of H.E.A.T*), Cash (*Cyborg 2: Glass Shadow*), Jessica (*Screamers*), Boomer (*Battlestar Galactica: The Plan*), The Machine (*The Machine*), and Ash⋆ (*Future World*). ⋆ Denotes characters with women chosen as their partners.

Figure 1: The *Maschinenmensch* before its transformation (*Metropolis*, 1927)



From its cinematic beginning, instances of the Lethal Seductress type are sexualized objects whose threat emanates from the presumption that females and femininity are benign, controllable and erotically alluring to men. Jeffrey Brown (2011: 101) confirms this, stating that 'the gynoid is always depicted as desirable despite being a living weapon hint[ing] at the depths of her value as a fetish'. This type's combination of desirability and violence is established by the inaugural instance of an F-PAWS in a live-action film: the *Maschinenmensch* (machine-man)[14] (portrayed by Brigitte Helm; see Figure 1) (*Metropolis*, 1927), created to replace the human Maria (also Helm), infiltrate the working classes, seduce upper-class men and destabilize the city. The *Maschinenmensch*'s performance is emotionally convincing to the workers and erotically irresistible to the gentry (see Figure 2), presenting the effective threat of the feminine mind and body when weaponized as an AWS.

Its ability to sway the workers to riot underscores the effectiveness of F-PAWS at infiltrating enemy spaces, subverting populations and toppling systems from within. Similarly, F-PAWS could function as seductresses,

[14] Here the German is using 'man' to indicate 'humanity' rather than the male sex, as the machine-appearance of the creation is gendered female and is referred to as such by its creator.

Figure 2: *Maschinenmensch*-Maria as the 'Whore of Babylon'

manipulating targets, gathering intelligence or executing covert assassinations within otherwise unassailable spaces. In either case, human operatives remain unexposed to risk. The horror instigated by an erotic machine stems from its manipulation of cisgender, heterosexual men's desires, ultimately requiring the gynoid's destruction because of her *sexual* threat (Short, 2011: 98). Thus, the *Maschinenmensch* is tied to a stake and burned alive, its Maria-facade contorting in manic laughter before melting away, revealing its metallic body. The *Maschinenmensch*'s weaponized femininity is sinister because it is controlled by Rotwang (played by Rudolf Klein-Rogge), who programmes the gynoid to destroy the city, while he watches from a safe distance. Rather than embodying the menace of uncontrolled F-PAWS, this instance underscores the peril of weaponized femininity directed at male society by an individual, specifically another man. The *Maschinenmensch* highlights a human-in-the-loop scenario because Rotwang programmes it to autonomously execute his destructive desires, exemplifying the dangers of human-passing AWS.

Metropolis, a German social allegory released in 1927 on 13 March (just months before the 3rd Nazi Party Congress was held in August of the same year), primarily seeks to convey a message of hope, presenting unity as the solution to a socially and economically crippled Germany. Nevertheless, its depiction of a successful infiltration and mass manipulation presents the dangers of a human-passing AWS when controlled by an individual intent on destabilizing the state. Additionally, this film cemented the trend of F-PAWS as infiltrators through the employment of their femininity.

The 1960s *Dr. Goldfoot* films[15] illustrate this concept through the Bikini Machines[16] and Girl Bomb[17] gynoids, which are tailored for specific targets. The Bikini Machines, printed from a machine, wear gold bikinis and follow Goldfoot's (played by Vincent Price) commands exclusively. They showcase AWS's functionality of autonomously executing directives and performing complex tasks to achieve programmed goals. When assigned a target, a persona and knowledge are downloaded into the machine's head, they don a trenchcoat to hide their bikini, then go to seduce and marry rich cisgender, heterosexual men, thus securing fortunes for Goldfoot. Alternatively, his Girl Bombs, identical to the Bikini Machines, are bespoke assassins that seduce and kill the targets Goldfoot has been paid to eliminate. Stuart Hall (2013: 256) states that fetishism involves 'the substitution of an "object" for some dangerous and powerful but forbidden force' and, in the case of F-PAWS, this 'forbidden force' is feminine violence, here substituted with fetishized gynoids. Purposefully created by their male controllers, their erotic appearances and actions feed into their fetishization. This recurring trend highlights the male controllers' weaponization of femininity to conceal their F-PAWS' threat. These gynoids are the gendered, highly advanced counterparts of AWS programmed for specific enemy types. They autonomously make rapid, high-level decisions to emulate human behaviour, acquire targets and fulfil their programmed objectives.

These 1960s films also began the trend of F-PAWS characters being 'excellent killers, [and] expendable' (Markowitz, 2019: 27) This expendability is rooted in their role as antagonists and the inherent violence associated with their portrayals. Additionally, it is a result of their representation as *things*, tools employed rather than individuals to be valued or mourned. This recurring theme of destruction and controllability is used by SF films to emphasize the F-PAWS' threat and required elimination. The depictions of these characters align with those of women in action and revenge films. However, as Sherrie Inness (1999: 81–82) states, action 'films also show that [killer] women are too tough and masculine', traits women should avoid or 'they will be punished, like the women in the films'. Notably, the *human* women Inness describes are *masculinized* by their violence or appearance (usually pronounced musculature), with no mention of them weaponizing their femininity. Alternatively, F-PAWS present an exaggeration of *femininity*, though they are occasionally masculinized when physically more capable

[15] *Dr. Goldfoot and the Bikini Machine* (1965) and *Dr. Goldfoot and the Girl Bombs* (1966).

[16] Played by: Patti Chandler, Mary Hughes, Salli Sachse, Luree Holmes, Sue Hamilton, Laura Nicholson, Marianne Gaba, China Lee, Issa Arnal, Deanna Lund, Pamela Rodgers, Leslie Summers, Sally Frei, Kay Michaels, Jan Watson, Arlene Charles and Alberta Nelson.

[17] Played by: Silvana Bacci, Antonietta Fiorito, Susan Hart and other uncredited performers.

or robust than an average woman. Consequently, these films introduce a different kind of 'killer women': inorganic substitutes controlled and weaponized by heterosexual men (Inness, 1999: 81).

Crucially, instances of the Lethal Seductress type depict F-PAWS as controlled things, advanced systems not responsible for their actions, merely executing pre-programmed directives. Therefore, these characters circumvent gender transgression; unlike the killer women in the action films Inness addresses, these gynoids are objects, not agents, tough only because *men* made them that way, control them, give them directives and a female presentation all to facilitate their infiltration and target-acquisition capabilities appealing to the expectation that *men* are the ones in control of combat and violence. These F-PAWs are weapons used, not persons acting.

However, some instances *are* sentient and act in that way, defying their controllers. The first F-PAWS to reject her creator-controllers' commands, making herself the first human out of the loop F-PAWS, is Andrea (played by Remy O'Neill) from *The Angel of H.E.A.T.* (1983). This film revisits the earlier trends such as the male creator employing his AWS army of gynoids *and* androids[18] to conquer the world. However, when Andrea is sent to infiltrate Agent Angel Harmony's (played by Marilyn Chambers) group, she falls in love with Samantha (played by Mary Woronov) Ultimately, Andrea betrays her creator, sacrificing herself to destroy his army and thus save Samantha. Significantly, Andrea marks the first gynoid in live-action cinema to choose a *woman* as her romantic partner. She also becomes the first F-PAWS to directly defy her creator and assert her agency, despite this resulting in her destruction. This depiction introduces a new trend of agentic F-PAWS which evolves so that by 2023, some characters survive similar circumstances.[19] Andrea's representation plants the seed within SF's fabular space that female-presenting weapons can have agency and will choose love, and that love will give them the strength to reject their lethal use. Yet, this makes Andrea's character a cautionary tale *against* gendering an AWS as female because they might be compromised by emotion and sexual attraction. Moreover, her defiance is predicated on love, accentuating the association between femininity and emotionality.

Notably, the 1990s relocate the origin of F-PAWS manufacture to corporate or military-linked sources, leading to a transition from farce and satire to more serious depictions of lethal creations. This shift in origin mirrors the reality of offscreen AWS's more prevalent existence by that time, like the Phalanx (General Dynamics, 1978) close-in weapons system used by

[18] Male-presenting instances with physical forms.

[19] For similar deaths, see *Screamers* (1996), *Battlestar Galactica: Razor* (2007) and *Battlestar Galactica: The Plan* (2009). For survivors, see *Tie Jia Wu Di Ma Li Ya* [*I Love Maria*] (1988).

the US military since 1980. Anna Gething (2011: 220) contends that violent women embody a combination of 'danger [and] desirability' reminiscent of the femme fatale trope popularized by film noir. As Katie Jones (2016: 36) observes, a femme fatale 'uses her sexuality against men', resulting in 'the necessity to control the feminine'. F-PAWS epitomize controlled femininity because their creators and controllers weaponize their danger and desirability. However, post-1980, instances of humans out of the loop increased as these lethal creations exceeded or lacked controllers. Brown (2011: 101) asserts that gynoids' 'value as a fetish' occurs because their objectification facilitates their violence. Thus, Gething and Brown highlight how these characters are sexualized fetishizations whose violence acts as fabular pleasure fantasies rather than presenting as agentic or menacing. Instances like Eve VIII (played by Renée Soutendijk in *Eve of Destruction*, 1991), Gaily (played by Clare Wren in *Steel and Lace*, 1991), Cash (played by Angelina Jolie in *Cyborg 2: Glass Shadow*, 1993) and Vanessa (played by Elizabeth Hurley in *Austin Powers: The Spy Who Shagged Me*, 1999) exemplify infiltrator-gynoids employing their female form to facilitate this goal while controlled, at least initially, by men.[20] These characters reflect SF's perception of women as nonthreatening to men, objects for sexual pleasure whose femininity is weaponized, thus reinforcing the expectation that human-passing F-PAWS are effective infiltrators.

The 1990s F-PAWS are deliberately depicted with erotic appearances and performances, aligning with the 1960s trend of F-PAWS as infiltrators, only now with military applications. The depictions of these characters present how female-presenting weapons are only effective if they are fetishized seductresses appealing to heterosexual men's desires. For example, Ms Connors (played by Pam Grier in *Class of 1999*, 1990), the first *named* gynoid of colour in live-action films, is a military F-PAWS engineered for infiltration, using her femininity to conceal her internal arsenal of weapons. When she is riddled with bullets and stabbed in the abdomen during the film's climax, she laughs and pulls open the flesh of her belly up to her clavicles, revealing her inner metal workings but also her naked artificial breasts. While the film's two AWS androids are similarly damaged, neither repeats this exposure, only peeling the flesh from their arms to release the weapons concealed therein. The emphasis on *her* nakedness accentuated through close-ups of her exposed torso and limbs, and her earlier pencil skirts and cleavage-revealing blazers, is juxtaposed with the clothed, concealing

[20] While there are instances where human women are similarly presented, such as the multiple Nikita adaptations – *La Femme Nikita* (1990), *Black Cat* (1991), *Point of No Return* (1993), *La Femme Nikita* (1997–2001), *Nikita* (2010–2013)), and series like *Alias* (2001–2006) or *Dollhouse* (2009–2010) – these depictions fall outside my scope of analysis.

dress of the androids throughout the film. Phrases like 'Here I am boys!' spoken during the final fight, coupled with her lascivious tone and attire, further suggest a distinct functionality from the androids. This depiction underscores the notion that female-presenting weapons are designed to be fetishized and appeal to the desires of heterosexual men, reinforcing the idea that seduction is integral for their effectiveness as infiltrators.

Continuing this military-made trend, Eve VIII (*Eve of Destruction*) is a military gynoid with a womb bomb who, like Ms Connors, is business-dressed, but more unassuming and awkward and only becomes seductive after she is damaged. This damage causes Eve VIII to exceed her programming, put on a leather, 'looking-to-get-laid', outfit and go on a man-harming spree. Alternatively, Gaily (*Steel and Lace*), like Ms Connors, is a walking arsenal, but has a nonmilitary creator and is the last instance of an F-PAWS with this origin. However, Gaily's embodiment is akin to the Bikini Machines and Girl Bombs: a perfect, walking fetishization for each of her targets, and then when the moment is right, she kills them. Later, Jessica (played by Jennifer Rubin in *Screamers,* 1996) is not fetishized in the traditional sense because she is not tailored for a specific target, but presents as a woman in a world with only cisgender, heterosexual men, so she easily infiltrates the human bunkers because of her perceived gender. It is worth noting that Jessica is machine-made; thus, even machines know that cisgender, heterosexual men will not view a woman as a potential threat. Each of these characters shows how their presented femininity is used to deceive cisgender, heterosexual men and enable the F-PAWS to achieve her mission of infiltration and/or assassination.

The trend of fetishized eroticization and infiltration continues in 21st-century films.[21] However, the last instance of femininity used as an infiltration tool within the scope of this chapter occurs in Alice's (played by Isabel Lucas in *Transformers: Revenge of the Fallen*, 2009) predatory, aggressive seduction of Sam (played by Shia LaBeouf). She repeatedly imposes herself on Sam, each instance visually presenting the threat of her sexual assertiveness on his feeble attempts to protest and resist his clear desire for her. Alice's sexually predatory depiction uses her sexual allure to get close to Sam, but only so she can kill him, suggesting the danger of female sexual deviancy to men.[22] She epitomizes the Lethal Seductress type by using erotic promise to successfully conceal her violent capability. This marks a significant shift as the first fully

[21] *Austin Powers in Goldmember* (2002), *Terminator 3: Rise of the Machines* (2003), *Battlestar Galactica: Razor* (2007) and *Transformers: Revenge of the Fallen* (2009).

[22] Supporting Deborah Tolman's and Tracy Higgins' (1996: 205) assertion that 'when women act as sexual agents, expressing their own sexual desire … they are often portrayed as threatening, deviant, and bad'.

autonomous, human-out-of-the-loop character while presenting a different kind of horror: AWS exceeding our control, addressed further in the next two sections.

3. Threatening Machines: unassuming, lethal femininity

The second type within the scope of this chapter is Threatening Machines: nonseductresses whose femininity is a deliberate choice by their creators for aesthetic appeal, infiltration or necessity for their designated function. These characters are ignored, underestimated or blend in *because* they present as female and thus are not perceived as threats by the humans in their films. This unassuming femininity presents the strength of gendered social expectations and interactions through the effectiveness of their feminine disguise. The first instance of this type is Host[23]-Tracy in the 1976 film, *Futureworld*. Host-Tracy (played by Blythe Danner) is created by other Hosts to replace her human precursor, a reporter, and infiltrate the human world to prepare for other human replacements until they are all replaced by Hosts. To do this, she must kill Tracy, but Host-Tracy fails to win their shootout despite being a complete copy of Tracy's mind and body. The film does nothing to explain how the human Tracy and Chuck (played by Peter Fonda) defeat their perfect artificial copies. Instead, it sets the stage for depictions of AWS weaponizing AWS against humans and failing, especially against human women, furthering this understanding of humans as the ones in control. It presents a reassuring fable of technological fallibility and machine logic failing to defeat human emotion and determination.

The Guardian (*Starcrash*, 1979) likewise fails in its function as the protector of the islands of the Amazons. Notably, *Starcrash* presents the rare instance of *female* controllers, the Amazons,[24] summoning the Guardian to protect their island from unwanted visitors. However, the Guardian's failure to defeat Stella (played by Caroline Munro) while under the Amazon's control and the Amazon's failure as combatants within the film, suggests that they are just fallible controllers overall. The fact that both Host-Tracy and the Guardian are defeated indicates that F-PAWS, while dangerous, can be stopped even by human women, a depiction repeated in the 1990s *Eve of Destruction* and *Steel and Lace*, and later in the *Resident Evil* films (2002, 2004, 2007, 2010,

[23] The name for the human-passing robots that are created by humans to populate pleasure parks that people pay to visit. The Hosts are there to allow humans the freedom to enact their fantasies without consequence.

[24] Played by Hélène Chauvin, Dirce Funari and Cindy Leadbetter; the rest are uncredited.

2012 and 2016) where Alice (played by Milla Jovovich) thwarts the Red Queen's[25] various attempts to kill her and humanity.

Unique within the Threatening Machines type are the F-PAWS with fetishistic embodiments contradictory to, or at least unnecessary for, their lethal function. These instances have an exaggerated femininity superfluous to their intended functionality, which says more about the fetishized female form in cinema than any perceived threat of femininity. While the hypermasculinized, muscle-man depictions of Arnold Schwarzenegger's various Terminator characters in the *Terminator* series (1984–2019) use exaggerated masculinity to present the strength of the character, the femininity of these Threatening Machine serves only as visual titillation for the viewer, with little to no comment upon it within the films.

In *Alienator* (1990), the hunter unit/Alienator (Teagan Clive) is a female-presenting and human-passing gynoid sent to hunt an escaped prisoner. Her performer is a bodybuilder, so her physical form is that of a muscular woman, yet she wears a metal bikini-like outfit and has white, heavy-metal hair. Significantly within this type, she does not attempt to blend in, nor does she use infiltration tactics to locate her target. Instead, she stands out amid the citizens of the unspecified American suburb who are all dressed in 1980s contemporary clothes. The film's poster foregrounds her image and that she is a 'woman' despite her being a machine within the film. The Alienator's human appearance does not fit the contemporary time on Earth, she makes no attempt to pass as human, nor does she use her sexualized dress as a way of seducing or enticing her target or anyone else. Thus, her appearance is ridiculous, an exploitative fetish there for viewers only, and by doing so negates the potential for her to be perceived as scary by them. The Alienator's depiction supports Tanya Krzywinska and Geoff King's (2006: 75) assertion that women's costumes in SF are 'physically revealing, [and] blatantly fetishistic'. Yvonne Tasker (1993: 19) also notes how filmmakers, responding to feminism, 'present women as active and as powerful' by using existing tropes like 'the leather-clad dominatrix' to present feminine power while maintaining their male viewers through fetishistic appearances. Thus, these fetishized, powerful women, as Brown (2011: 73) notes, 'cater to [this] very specific sexual fantasy', making these F-PAWS into sexual fantasy figures: fetishizations for their viewers rather than dangerous threats despite their clear, violent capability. Additionally, it is worth noting that the Alienator is the first instance of an F-PAWS having differing pronoun use, as her controllers refer to her as 'it', while the humans

[25] Played by Michaela Dicker in *Resident Evil* (2002), body: Megan Charpentier voice: Ave Merson-O'Brian in *Resident Evil: Retribution* (2012) and Ever Gabo Anderson in *Resident Evil: The Final Chapter* (2016).

on Earth all refer to her as 'she'. The fact that the Alienator and most of the other F-PAWS within the scope of this chapter of films are referred to with female pronouns by the characters who do and do not know they are artificial reflects the strong link between presented gender and perceived sex.

The fetishized embodiment trend within the Threatening Machines type diminishes notably by the late 2000s, as F-PAWS characters increasingly appear as sentient agents rather than titillating weapons. Their violence remains prominent, but they are sentient agents whose femininity is not erotically emphasized. Examples of this are AMEE (*Red Planet*, 2000), the Arcee sisters: Arcee (voice: Grey Griffin, rider: Erin Naas), Chromia, and Elita (*Transformers: Revenge of the Fallen*), and Shatter (voiced by Angela Bassett) and Arcee (voiced by Grey Griffin in *Bumblebee,* 2018), all of whom are robots who are machine-appearing, although the Transformer characters are humanoid in shape. Their lack of a human appearance means they do not rely on eroticization. Instead, these machine-presenting characters are feminine through voice, shape, colouration or name, and their violence is more akin to male-presenting instances where their gender does not play a part in their functionality. This subset of machine-appearing characters,[26] 11 out of the over 57 F-PAWS characters in live-action films, counters Brown's (2011: 101) assertion that 'the gynoid is always depicted as desirable despite being a living weapon', since there are some that are not. Yet their minority speaks to the prevalence of F-PAWS' fetishization and erotic depictions. These machine-appearing gynoids exemplify a transition away from fetishization by emphasizing their strength and lethal capacity while still being perceivably feminine and eschewing overt sexual allure.

An alternative trend in the late 1990s that continues into the 2000s is the growing prevalence of F-PAWS originating from the Military-Industrial Complex (MIC) reflecting the influence of real-world militaries' adoption of AWS technology.[27] These depictions are cautionary tales highlighting the dangers of granting increased control to autonomous AWS systems without proper human oversight. Instances such as the AI ARIIA (voiced by Julianne Moore in *Eagle Eye*, 2008) and the *Terminator* series' Skynet exemplify how advanced military AI systems designed to enhance security could surpass human control and attempt to seize control. This loss of control is not exclusive to female-presenting AI, as male-presenting AI

[26] Guardian (*Starcrash*), AMEE (*Red Planet*), Arcee, Chromia and Elita (*Transformers: Revenge of the Fallen*), Beauty-Bot (*Kingsman: The Golden Circle*, 2017), Quintessa (Gemma Chan, *Transformers: The Last Knight*, 2017), Shatter and Arcee (*Bumblebee*, 2018), Mother (body: Luke Hawker, voice: Rose Byrne, *I Am Mother*, 2019) and Robot Bubs (*Seungriho*).

[27] See Opfer, 2014; McFadden, 2018; and Global Defence Technology, 2018.

and robots also demonstrate similar behaviour.[28] However, these AI pose a more credible threat due to their similarity to existing AWS and AI in use or being developed by the MIC (see Caton, 2015). This convergence of onscreen and offscreen technologies again reflects how SF explores potential realities through horror and comforting fables to warn or console us about our technological futures.

Alternatively, in 21st-century films, AI characters are neither seductresses nor superfluously eroticized. Instead, these AIs present human-like holograms and voices designed to be comforting and initially nonthreatening, with their lethal abilities and intentions revealed later on. This shift away from fetishization speaks to a broader trend in SF films, aligning with offscreen AWS and AI that lack overt fetishization and feminization. For example, the Red Queen from the *Resident Evil* films presents a hologram-child interface, evoking suspicion and eerie horror by juxtaposing innocence with her capacity for violence, as she tells the humans: 'You're all going to die down here.' Importantly, while threatening because of her projection's colour, actions and words, she is never eroticized in her depiction.

Unlike the military-made AI of the new millennium, the live-action gynoids from the 2000s are predominantly created by sentient machines to harm humans overall or other machines rather than being tailored to appeal to men. This shift in origin grants these characters a level of autonomy detached from human control. Like the Transformers, these characters are made by machines, possess agency from the outset and exist outside human control, cooperating with humans when it suits their interests. These instances continue Jessica's (*Screamers*) depiction as a killing machine made by other machines to infiltrate human-controlled spaces in order to eliminate them. While the Screamers and Terminators illustrate a progression from human creations to autonomous entities, the Transformers, being extraterrestrials, offer a unique perspective as artificial beings modelled after human machines to blend into society.[29] Their machine camouflage echoes other F-PAWS's female camouflage, which likewise allows them to blend in because women are not considered threatening. The shift from human-made origins to entirely extraterrestrial origins highlights SF's evolution towards autonomous, uncontrollable adversaries, distancing itself from its former eroticizations.

[28] See *Der Herr Der Welt* (1934), *Tobor the Great* (1954), *Kronos* (1957), *The Colossus of New York* (1958), *The Creation of the Humanoids* (1962), *Alphaville* (1965), *2001: A Space Odyssey* (1968), *Colossus: The Forbin Project* (1970), *Westworld* (1973), *Dark Star* (1974), *Futureworld* (1976), *Demon Seed* (1977), *Saturn 3* (1980), *Blade Runner* (1982), *Wargames* (1983), all the *Terminator* films, *Chopping Mall* (1986), *Deadly Friend* (1986), *Class of 1999* and *Hardware* (1990), to name but a few.

[29] They evolved on their planet Cybertron on their own, their wars and conflicts now playing out on Earth because they destroyed Cybertron during their wars.

Moreover, the diminishing instances of F-PAWS as infiltrators and assassins underscore a transition towards overtly depicted female-presenting soldiers. This transformation reflects the US MIC's substantial investment in AWS, robotics, AI and other artificial assistive technologies, all aimed at automating military operations, with machines taking active roles on the battlefield, enhancing soldiers, and maintaining martial technological dominance.

4. Uncontrolled Agents: fully autonomous lethal femininity

The final type – Uncontrolled Agents – are AWS who are no longer controlled or never were controlled. As noted in the Threatening Machines type, towards the end of 1999, films began to present F-PAWS' lethality without fetishization and more as overt threats. In *Omega Doom* (1996), the film's Roms,[30] an all-female-presenting group, have near-identical costumes and appearances: Caucasian skin tones, severe, short, black, straight hair, black sunglasses, and black pleather-like outfits of high-necked shirts, trench coats, trousers and combat boots. Their voices are modulated so that their female voices have a deeper sound to them. Their eyes are metallic orbs, which makes them menacing and uncanny in appearance compared to the other artificials of the film, which have human-appearing eyes. These characters are the most advanced models in the film's world, created to be killing machines before the manufacture of artificials stopped. They are soldiers only and understand little else, except for Zinc (played by Jill Pierce), who, in a moment of honesty, says that she never enjoyed killing, and Bartender (played by Anna Katarina) sees Zinc's now human-appearing eyes, signalling that she is not a perfect soldier after all. *Omega Doom*'s presentations suggest that F-PAWS' onscreen representations did change, albeit slowly, over the 20th century.

An example of this shift is Quintessa (played by Gemma Chan in *Transformers: The Last Knight*, 2017), who offers a striking departure from previous instances of F-PAWS by dominating male characters through force alone. Unlike previous antagonists within the *Transformer* film series who rely on size and physical dominance, Quintessa defeats the much larger Optimus Prime (voiced by Peter Cullen) with a wave of her hand. Optimus is a humanoid machine the size of a three-storey house, so his imposing stature versus her diminutive one (she could fit in his hand) positions him as the expected victor. Instead, she proclaims herself 'the Prime of Life',

[30] Blackheart (Tina Cote), Zinc (Jill Pierce), Ironface (Cynthia Ireland), unnamed: Silvia Matasovska, Katarina Borova, Ingrid Hazalovicova, Jana Korcova, Denisa Capkova, Katka Alexejevicova, Ivana Andrejkovicova, Jarmila Pejchlova, Jurj Dulik and Ivan Lorenc.

a god to the Transformer race, and acts accordingly. She commands the other Transformers and when she physically marks Optimus with a slap, she shows her dominance over him. This action metaphorically 'castrates' him, stripping away his weaponry and free will by using her Prime abilities to reprogramme him into Nemesis Prime.[31] In this way she embodies uncontrolled femininity: untameable, unconquerable and uninterested in male figures or society, except in terms of dominance. This domination scene underscores Quintessa's role as a formidable feminine force that can 'emasculate' male autonomy and authority, replacing it with submission to her will. Quintessa's close-ups show an attractive (albeit metal) face and form, but do not linger or use extreme close-up shots to accentuate her form. She has the most human-appearing Transformer form, signalling her function as a cerebral force rather than the other Transformers' physical and mechanical appearances. Her representation and that of the other gynoids within the *Transformer*s film series signal the change in F-PAWS' representational trends towards female-presenting castration threats: autonomous, agentic soldiers whose violent abilities and commanding authority pose a serious, uncontrolled threat to men (Brown, 2011: 101).

Alternatively, the last two films within the analysed scope demonstrate a distinct shift in the Uncontrolled Agents type. *I Am Mother* (2019) presents a future where an AI takes over Earth, almost eradicating humanity, because it concludes that humans will destroy themselves. It creates a bunker with stored embryos and an ancillary robot body (performed by Luke Hawker and voiced by Rose Byrne). This ancillary, called 'Mother', is designed for rearing human children, equipped with gel pads for infant care, combined with a metal body for protecting Daughter (played by Clara Rugaard). Daughter is raised by Mother alone in the bunker, undergoing various tests to ensure that her physical health and moral integrity align with the AI's ideal for its new humanity. The film concludes with a revelation: the AI allowed a human Woman (performed by Hilary Swank) into the bunker to test Daughter's loyalty to the AI or choose to join the Woman's defiance of the AI. Daughter ultimately chooses Mother, returning to her newly decanted brother and is told that Mother made the surface uninhabitable, then restored it for the forthcoming generation of humans she and Daughter will raise. This film presents an AI AWS that exceeds human expectations, yet is still following its directive to keep humanity safe, just by very unexpected means.

Significantly, Mother is physically played by a man and voiced by a woman. This is the first instance of a cross-gender performance of an F-PAWS character. By doing this, the production speaks to the performative

[31] The name for Optimus after Quintessa dominates him, overriding his own will and making him fight for her.

nature of gender by having the voice, function and behaviour of Mother be the only thing that sets her apart from the other ancillary forms of AI. In this sense, the humans perceive and interact with her as if she is female, while truly she, as an ancillary of an AI that is controlling all the robots and machines on the planet, is without gender. The AI's robot form was inspired by Boston Dynamics' Atlas (2013) humanoid robot (Evangelista, 2019). This design and performance choice speaks to how contemporary technologies can affect SF and how SF perceives women as more nurturing, despite the robot's masculine frame. Additionally, it shows how, once sentient, an AWS can *choose* their gendering, though Mother's gendering is more a function of femininity's association with childrearing than an aspect of their identity, since the AI's countless other ancillary humanoid bodies are identical to Mother, only without the nurturing components, while its other small drones and the larger farming machines have functional forms.

The last instance of an F-PAWS of this type is Robot Bubs (voiced and performed by Yoo Hae-Jin in *Seungriho* [*Space Sweepers*], 2021), who does not outwardly present as female until the end of the film. Bubs' instance is a military-made robot designed to be a soldier who now works with a crew of space sweepers who collect space debris for money. Significantly, Bubs is an equal part of the crew, getting her cut of their pay and helping harpoon debris because she can be outside the spaceship. She is not controlled by anyone, but is an equal onboard the ship. While on the various stations or other ships, she will pretend to be a standard, controlled robot, but her crew knows otherwise, as do her friends on the other scrapper ships. This male-voiced, machine-appearing artificial creation is sentient, has agency and desires, and a sense of personal identity. Bubs also identifies as female despite her appearance and programming lacking any nuances of gender beyond her built-in male voice. Bubs' instance of an artificial creation having agency – and an identity that differs from the one it was created with – is a first for AWS in live-action films, and for AI and robots within live-action films overall.[32]

At the film's end, Bubs is fully human-passing, having purchased a skin graft so she can pass as a woman (played by Kim Hyang-Gi). However, she still has her male voice because she has not purchased a female one yet. The viewer does not know that Bubs identifies as female until the middle of the film when Dorthey/Kang Kot-Nim (played by Park Ye-Rin), a little girl

[32] The only other instance where a robot or AI's presented gender changes within live-action films occurs in *Rocky IV* (1985), where Sico (male voice: Robert Doornick, female voice: uncredited) has its voice and behaviour changed by Paulie (Burt Young), its owner, because he is unnerved by the male, robotic voice and so changes it to a female one which is demure and adoring of him.

with nanobots that allow her to interface with technology, is sitting with Bubs while Bubs does her makeup and Dorthey calls Bubs 'Unnie' (a Korean term used by a younger woman when speaking to an older woman). When Bubs hears this, she gets flustered and even has red blush-lights appear on her faceplate, then confesses to Dorthey that she is 'thinking of getting skin grafts', but is 'scared people will laugh' at her. It is only after listing her former military functions that she finally realizes that Dorthey is not a robot as well.[33] This exchange presents not only how Bubs was formerly used as a weapon with no regard for her agency, but also how she was discarded once her usefulness ended (we are told the ship's captain found her in a recycling centre). This expendability echoes that seen in earlier films; however, Bubs' agency and uncontrolled state enable her to get a new chance at living life on her terms rather than just as a weapon.

Both Mother and Bubs exemplify the culmination of the F-PAWS' trends within live-action films up to 2023. Initially created by humans as AWS for various purposes, these characters surpass their intended roles and demonstrate agency, defiance and independence. Mother follows her directives in unexpected ways, while Bubs embodies the desire for a more human experience by *choosing* to be human-passing and female, contrary to her initial programming. Both characters are created without female presentations, but later create or choose female forms and presentations, showing the level of their agency. These characters deliberately choose their female presentations rather than having them imposed to serve male-centric roles, marking a shift towards greater agency and empowerment. This evolution underscores the changes within the F-PAWS archetype beyond mere seductresses, infiltrators or submissive entities, towards characters with multifaceted agency and personal aspirations while cementing the performative nature of gender through their choice of presentation.

Conclusion

The initial depictions of F-PAWS emphasize their controllability and fetishization, positioning them as expendable commodities manipulated by men for destructive purposes. This disposability aligns with AWS' offscreen role as substitutes for humans in combat scenarios (Caton, 2015; Etzioni, 2018). The point of military AWS is to be a 'force multiplier' so that 'fewer soldiers are needed for a given mission and the efficacy of each soldier is greater', to 'expand the battlefield, allowing combat to reach into areas that

[33] The crew believed she was a walking bomb because of her wanted advertisements labelled her as such as a means of ensuring people would report her whereabouts and the corporate military could pick her up.

were previously inaccessible' and to 'reduce casualties by removing human soldiers from dangerous missions' (Etzioni, 2018: 253). F-PAWs achieve these three goals by replacing soldiers, thus reducing their number and sparing soldier casualties, but especially through their female-presentation, which grants them access and allows them to infiltrate targeted areas and persons. F-PAWS, like offscreen AWS, are controlled by men, but if not destroyed in their film during this period, they remain subservient to male characters, functioning as tools and objects for heterosexual men either way rather than being autonomous threats. These depictions speak to the expectation for such creations as obedient to *men*, weaponized by men against men.

As subsequent F-PAWS continue this trend of specialist infiltrators or assassins, they do so without seductive tactics. This evolution continues the AWS-as-infiltrator portrayal while moving towards being autonomous soldiers. This transition signifies a cinematic shift towards agentic and sentient depictions. These later F-PAWS characters are cautionary figures, warning of the repercussions of treating sentient entities as controlled tools and showing what could happen if AWS are fully autonomous and programmed with higher-level thinking. Furthermore, they explore how the feminine form, associated with comfort rather than threat, enables F-PAWS to infiltrate and manipulate heterosexual men while simultaneously leading adversaries to underestimate them. Yet, these same depictions often lose to humans, presenting a reassuring tale of human dominance over our technologies and over independent women. For instances that are alien in origin, there is a new shift that shows sentient machines fighting against other sentient machines and humanity. These depictions do not require their F-PAWS to infiltrate; instead, they are soldiers fighting for a cause. These instances present an ideal AWS that would be a fully autonomous soldier fighting for us against enemies both mechanical and organic.

Significantly, recent instances like Mother and Robot Bubs demonstrate that autonomy does not equate to nonviolence; rather, it grants them the ability to *choose* when and whom to fight. This level of independent decision making is a primary concern for policy makers, humanitarians and the military, as the loss of control raises questions about safety and the feasibility of regaining authority or terminating these weapons. Similarly, the cinematic emergence of AI in the 2000s parallels the rise in offscreen AI technologies and virtual assistants, and these filmic AI's insidiousness taps into and magnifies the anxieties we experience whenever our devices fail to obey us. Thus, the prospect of rogue AI, exemplified by ARIIA and Skynet, resonates strongly with contemporary audiences because of their ability to infiltrate and manipulate all electronic devices.

The rise in villainous F-PAWS characters mirrors US society's (and the world's) unease regarding uncontrollable machines, stemming from the increasing prevalence of drone warfare, military demand for AWS and relentless advancements in robotics geared towards creating human-passing robots (Biba, 2022). These 43 films envision potential futures for AWS, robots, AI and our social understanding of femininity and its places within conflict. While some films offer reassurance through humans triumphing over our out-of-control creations, others present the (horror) story of what can happen when AWS take control. Most often, these tales begin with a military origin and speak to the dangers and needlessness of violence, especially once it becomes a matter of AWS against other AWS removing human soldiers from the battlefield while showing the toll it takes on civilians and infrastructure. Significantly, more recent depictions show AWS working with humanity, seeking to protect, nurture and lead them towards a better future, promoting peace over escalating war, advocating against AWS's offscreen use and existence, and showing how SF explores alternate technological futures.

Filmography

Alienator (1990) Directed by Fred Olen Ray. US: Amazing Movies.

Android Apocalypse (2006) Directed by Paul Ziller. US and Canada: Independent Moving Productions Inc.

Angel of H.E.A.T. (1983) Directed by Myrl A. Schreibman. US: Levy Films.

Austin Powers in Goldmember (2002) Directed by Jay Roach. US: New Line Cinema.

Austin Powers: International Man of Mystery (1997) Directed by Jay Roach. US and Germany: Capella International.

Austin Powers: The Spy Who Shagged Me (1999) Directed by Jay Roach. US: New Line Cinema.

Battlestar Galactica: Blood & Chrome (2012) Directed by Jonas Pate. US: Universal Pictures.

Battlestar Galactica: Razor (2007) Directed by Félix Alcalá. US: Universal Pictures.

Battlestar Galactica: The Plan (2009) Directed by Edward James Olmos. US: Universal Pictures.

Bumblebee (2018) Directed by Travis Knight. USA & China: Paramount Pictures.

Class of 1999 (1990) Directed by Mark L. Lester. US: Lions Gate Home Entertainment.

Cyborg 2: Glass Shadow (1993) Directed by Michael Schroeder. US: Anglo-American Film Corporation.

Dr. Goldfoot and the Bikini Machine (1965) Directed by Norman Taurog. US: Alta Vista Productions.

Dr. Goldfoot and the Girl Bombs (1966) Directed by Mario Bava. US and Italy: American International Pictures.

Eagle Eye (2008) Directed by D. J. Caruso. US and Germany: DreamWorks.

Eve of Destruction (1991) Directed by Duncan Gibbins. US: Nelson Entertainment.

Future World (2018) Directed by Bruce Thierry Bheung and James Franco. US: Lionsgate.

Futureworld (1976) Directed by Richard T. Heffron. US: American International Pictures.

Ghost in the Shell (2017) Directed by Rupert Sanders. US, India, Hong Kong and China, Canada: Paramount Home Media Distribution.

I Am Mother (2019) Directed by Grant Sputore. Australia: Southern Light Films.

Kingsman: The Golden Circle (2017) Directed by Matthew Vaughn. UK and US: 20th Century Fox.

M3GAN (2023) Directed by Gerard Johnstone. US and New Zealand: Atomic Monster, Blumhouse Productions, Divide/Conquer.

Metropolis (1927) Directed by Fritz Lang. Germany: Universum Film.

Oblivion (2013) Directed by Joseph Kosinski. US: Universal Pictures.

Omega Doom (1996) Directed by Albert Pyun. US: Columbia TriStar Home Video.

Red Planet (2000) Directed by Antony Hoffman. US and Australia: Warner Bros.

Resident Evil (2002) Directed by Paul W.S. Anderson. UK, Germany, France and US: Constantin Film.

Resident Evil: Retribution (2012) Directed by Paul W.S. Anderson. Germany, Canada, US, France and UK: Constantin Film International.

Resident Evil: The Final Chapter (2016) Directed by Paul W. S.Anderson. UK, France, Germany, South Africa, Canada, Japan, Australia and US: Constantin Film.

Screamers (1996) Directed by Christian Duguay. Canada, USA and Japan: Triumph Films.

Seungriho [*Space Sweepers*] (2021), Sung-hee Jo. South Korea: Bidangil Pictures.

Sky Captain and the World of Tomorrow (2004) Directed by Kerry Conran. UK, Italy and US: Paramount Pictures.

Some Girls Do (1969) Directed by Ralph Thomas. UK: Rank Organisation.

Starcrash (1979) Directed by Luigi Cozzi. US: Nat and Patrick Wachsberger Productions.

Steel and Lace (1991) Directed by Ernest D. Farino. US: Cinema Home Video.

Terminator 3: Rise of the Machines (2003) Directed by Jonathan Mostow. US, Germany and UK: C-2 Pictures.

Terminator Salvation (2009) Directed by McG. US, Germany, UK and Italy: Halcyon Company.

The Machine (2014) Directed by Caradog W. James. UK: Red & Black Films.

Tie Jia Wu Di Ma Li Ya [*I Love Maria*] (1988) Directed by David Chung, Hark Tsui. Hong Kong: Film Workshop.

Tomorrowland (2015) Directed by Brad Bird. US and Spain: Walt Disney Pictures.

Transformers: Revenge of the Fallen (2009) Directed by Michael Bay. US: DreamWorks.
Transformers: The Last Knight (2017) Directed by Michael Bay. US, China and Canada: Paramount Pictures.
Virtual Girl (1998) Directed by Richard Gabai. US: Check Entertainment Distribution.

References

Bhuta, N., Beck, S., Geiβ, R., Liu, H.-Y. and Kreβ, C. (eds) (2016). *Autonomous Weapons Systems: Law, Ethics, Policy*. Cambridge: Cambridge University Press.
Biba, J. (2022) 'Top 20 humanoid robots in use right now', *BuiltIn,* 14 September. Available from: https://builtin.com/robotics/humanoid-robots [Accessed 12 November 2022].
Brown, J.A. (2011) *Dangerous Curves: Action Heroines, Gender, Fetishism, and Popular Culture*, Jackson: University Press of Mississippi.
Butler, J. (2006) *Gender Trouble: Feminism and the Subversion of Identity*, Abingdon: Routledge.
Caton, J. (2015) *Autonomous Weapon Systems: A Brief Survey of Developmental, Operational, Legal, and Ethical Issues*, Carlisle, PA: US Army War College Press. Available from: https://press.armywarcollege.edu/monographs/304. [Accessed 11 November 2022].
Department of Defense (2017) *Directive 3000.09: Autonomy in Weapon Systems*, Washington DC: Department of Defense.
Etzioni, A. (2018) 'Pros and cons of autonomous weapons systems (with Oren Etzioni)', in A. Etzioni (ed) *Happiness Is the Wrong Metric: A Liberal Communitarian Response to Populism*, Cham: Springer International Publishing, pp 253–263.
Evangelista, C. (2019) 'Here's how Weta created the *I Am Mother* robot', */Film*, 12 June. Available from: https://www.slashfilm.com/567013/i-am-mother-featurette/ [Accessed 13 December 2023]
Gething, A. (2011) '"A Caligula-like despot": matriarchal tyranny in *The Sopranos*', in M. Waters (ed) *Women on Screen Feminism and Femininity in Visual Culture*, Basingstoke: Palgrave Macmillan, pp 213–224.
Global Defence Technology (2018) 'Robotic warfare: training exercise breaches the future of conflict', *Global Defence Technology*, November. Available from: https://defence.nridigital.com/global_defence_technology_special_nov18/robotic_warfare_training_exercise_breaches_the_future_of_conflict [Accessed 3 May 2019].
Hall, S. (2013) 'Spectacle of the "other"', in S. Hall, J. Evans and S. Nixon (eds) *Representation*, 2nd edn, Milton Keynes: Open University Press, pp 215–271.

Heyns, C. (2016) 'Autonomous weapons systems: living a dignified life and dying a dignified death', in N. Bhuta et al (eds) *Autonomous Weapons Systems: Law, Ethics, Policy*, Cambridge: Cambridge University Press, pp 3–19.

Inness, S.A. (1999) *Tough Girls: Women Warriors and Wonder Women in Popular Culture*, Philadelphia: University of Pennsylvania Press.

International Committee of the Red Cross (2014) 'Autonomous weapons systems: technical, military, legal and humanitarian aspects'. Available from: https://www.icrc.org/en/download/file/1707/4221-002-autonomous-weapons-systems-full-report.pdf [Accessed 15 November 2022].

International Committee of the Red Cross (2016) *Autonomous Weapons Systems Implications of Increasing Autonomy in the Critical Functions of Weapons*, Versoix: International Committee of the Red Cross (ICRC), 26–28 March. Available from: https://icrcndresourcecentre.org/wp-content/uploads/2017/11/4283_002_Autonomus-Weapon-Systems_WEB.pdf [Accessed 6 November 2022].

International Committee of the Red Cross (2021) 'Autonomous weapons: the ICRC recommends adopting new rules', 3 August. Available from: https://www.icrc.org/en/document/autonomous-weapons-icrc-recommends-new-rules [Accessed 15 November 2022].

Jones, K. (2016) 'Bluebeardean futures in Alex Garland's *Ex Machina* (2015)', *Gender Forum*, 58: 20–40.

Jones, R. (2022) 'Artificial sexuality & sentience: the evolving archetypal representations of female-presenting artificial intelligence and gynoids in live-action American and British film', unpublished doctoral thesis, De Montfort University, Leicester.

King, G. and Krzywinska, T. (2006) *Science Fiction Cinema: From Outerspace to Cyberspace*, London: Wallflower Press.

Liang, Y. (2015) 'Female body in the postmodern science fiction', *Theory and Practice in Language Studies*, 5(10): 2037. http://dx.doi.org/10.17507/tpls.0510.09

Markowitz, J.A. (2019) *Robots That Kill: Deadly Machines and Their Precursors in Myth, Folklore, Literature, Popular Culture and Reality*, Jefferson, NC: McFarland & Company, Inc.

McFadden, C. (2018) 'A brief history of military robots including autonomous systems', *Interesting Engineering,* 6 November. Available from: https://interestingengineering.com/innovation/a-brief-history-of-military-robots-including-autonomous-systems [Accessed 3 May 2019].

Opfer, C. (2014) 'Are robots replacing human soldiers?', *HowStuffWorks*, 13 February. Available from: https://science.howstuffworks.com/robots-replacing-soldiers.htm [Accessed 3 May 2019].

Short, S. (2011) *Cyborg Cinema*, Basingstoke: Palgrave Macmillan.

Siegel, M., Breazeal, C. and Norton, M. I. (2009). 'Persuasive Robotics: The Influence of Robot Gender On Human Behavior'. pp 2563–2568. Institute of Electrical and Electronics Engineers.

Strait, M.K., Aguillon, C., Contreras, V. and Garcia, N. (2017) "The public's perception of humanlike robots: online social commentary reflects an appearance-based uncanny valley, a general fear of a "technology takeover", and the unabashed sexualization of female-gendered robots", presented at the 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), IEEE, pp 1418–1423.

Tasker, Y. (1993) *Spectacular Bodies: Gender, Genre, and the Action Cinema*, London: Routledge.

Tolman, D.L. and Higgins, T.E. (1996) 'How being a good girl can be bad for girls', in N.B. Maglin and D.M. Perry (eds) *'Bad Girls'/'Good Girls': Women, Sex, and Power in the Nineties*, New Brunswick, NJ: Rutgers University Press, pp 205–225.

WELT (2012) 'Feature: von IBM Shoebox bis Siri: 50 Jahre Spracherkennung', *DIE WELT*, 20 April, Available from: welt.de/newsticker/dpa_nt/infoline_nt/computer_nt/article106206488/Von-IBM-Shoebox-bis-Siri-50-Jahre-Spracherkennung.html [Accessed 22 November 2022].

Wosk, J. (2015) *My Fair Ladies: Female Robots, Androids, and Other Artificial Eves*, New Brunswick, NJ: Rutgers University Press.

Autonomous Weapons in Fiction and the Fiction of Autonomous Weapons

Teresa Heffernan

Introduction

In March 2022, a Russian suicide drone, the KUB-BLA, which claims to use artificial intelligence (AI) to identify targets, was spotted in Ukraine. This small strike drone is an unmanned aerial vehicle that uses a visual identification system; it can hover above a target, carry 3 kg of sensors and explosives, travel up to 130 km/h and explode on impact (Army Technology, 2023). The use of this drone escalated concerns about the use of lethal autonomous weapon systems (LAWS) and 'the prospect that more capable systems could eventually decide for themselves who to kill' (Knight, 2022). The capacity to decide is one of the ways autonomous weapons are defined. But how can a machine decide – as in making up its own mind – when it does not have a mind?

Alan Turing speculated in 1950 that by the end of the 20th century, 'the use of words … will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted' (Turing, 1950: 442), and many AI researchers have followed suit, altering the meaning of words to accommodate machine logic. However, this 'altering' has distorted discussions of AI and LAWS, which use some combination of sensor suites and computer algorithms to collect and analyse data in order to identify and strike targets. Yet anthropomorphic rhetoric that credits machines with traits such as autonomous decision making and intelligence nevertheless frame debates about their use. As Bächle and Bareis argue, 'semantic ambiguities' in the debates and definitions of LAWS in China and the US, like the blurring of the distinction between autonomy and automatic, help to undermine any meaningful regulation and help countries to skirt the Geneva Convention as they develop 'highly automatic and destructive weaponry' (Bächle and Bareis, 2022: Conclusion section).

References to fiction further encourage the anthropomorphism and animation of inanimate machines. Repeating a common mantra, a *New York Times* article describes autonomous weapons as having 'jumped from the pages of science fiction to reality', while the Pentagon often refers to the 'Terminator conundrum' in the context of AI warfare (Rosenberg and Markoff, 2016). This representation of LAWS as originating in fiction and springing from its pages, amplified by the Terminator metaphor that has long pervaded media, policy and academic discussions, has not only shaped the reception and sociotechnical imaginaries of these weapons (Cave and Dihal, 2019; Watts and Bode, 2023), but has also distracted from their role in the expansion of the global arms trade. Modelled on the much-critiqued theory of the autonomous liberal subject, which fails to account for the ways in which humans extend into networks that exceed individual cognition, the labelling of weapons as autonomous similarly fails to account for the powerful interests shaping the development of this technology. Rather than animating the technology by way of fiction, we should be asking, as Lucy Suchman (2019: 36) does: 'In whose interests are these projects, and who decides that they should go forward, in lieu of other projects of transformative future making?'

Moreover, the prevalent fiction to reality framing of AI technology has short circuited the ethical power of fiction to question the ideological underpinnings of a world that produces them. From the first mention of robots in Karel Čapek's play *R.U.R.: Rossum's Universal Robots* (1920) to James Cameron's three *Terminator* films (1984, 1991 and 2019),[1] it is the human as a machine, not literal autonomous machines, that bring about the destruction of the world. However, the crucial insights these works of fiction offer, about how not to destroy the world, get lost when fiction as fiction is disavowed and fictional robots or cyborgs are conflated with real weapons. Reading fictional robots and cyborgs as literal as opposed to liminal figures that interrogate what it means to be human prevents us from taking responsibility for the worlds we create and fails to address the complex topographies of fiction that foreground figurative language.

The first section of this chapter considers LAWS not as autonomous machines that leap from the pages of fiction, but as extensions of the lucrative global arms trade, which often overrides the debates about their ethical use. The second section situates LAWS in the context of

[1] Cameron regretted selling the rights to *The Terminator* and regained them for *Dark Fate*, which ignores the storylines from the other films and continues on from *Terminator 2*. He was involved in the story for all three of these films.

the human as machine and historic ties between the military and the AI industry. The third section argues that reading the three James Cameron *Terminator* films as about a soon-to-be-realized weapon and the future of war misses the point of these films, which expose the techno-military-industrial complex and the myths circulated by the AI industry. The fourth section returns to the origins of killer robots in *R.U.R.* and the fantasy of humans as efficient profitable machines that culminates in a world-destroying war.

1. Lethal autonomous weapon systems and the arms industry: who are the 'bad guys'?

Debates about LAWS have focused on anything from the precision of sensors and algorithms to concerns about the ethics of using these weapons. The arguments for and against their use to locate, track, attack and kill human targets have been well rehearsed: those in favour of developing these weapons refer to such things as the advantage of their speed in decision making, the precision of their aim, the scale of their reach, their dispassionate objectivity, the fear that another country might gain the lead, the protection of human soldiers from harm and the reduction of military costs, as machines are argued to be cheaper than soldiers. Those against point to the fallibility of the technology, the violation of international humanitarian law, automation bias that encourages humans to not question machines, the escalation of conflict and a new arms race, the use of LAWS by nonstate actors and authoritarian regimes, the automation of war, and the ethical implications of and moral responsibility for LAWS that kill humans (Etzioni, 2018).

In March 2018, I co-hosted an event at Saint Mary's University in Halifax on LAWS: Noel Sharkey, a computer scientist at the University of Sheffield, chair of the International Committee for Robot Arms Control, and one of the founders of the Campaign to Stop Killer Robots, debated Duncan MacIntosh, a philosopher at Dalhousie University and advisor to several US think tanks. MacIntosh argued, among other things, that these weapons will allow the 'good guys' to more clearly target and efficiently kill the 'bad guys' as a preventative measure. Sharkey, in turn, demonstrated how easy it would be to trick the machine into focusing on the wrong target and pointed to the fallibility of these weapons. Beyond the crudeness of the technology, many students who attended the packed event objected to the default position of the debate that seemed to accept war as a constant and inevitable fact of modern life, bypassing more reasonable approaches to sorting through global conflicts. Unwilling to accept the Orwellian doublethink of 'War is Peace' as long as the target is far away and the carnage is out of sight, the students questioned the 'good guy' versus 'bad guy' argument, an argument that is further complicated when the lens is widened to include the business of weapons.

Following the campaigns mounted by activists and academics against 'killer robots', the United Nations (UN) has been debating the question of banning or at least regulating LAWS. At a meeting in Geneva in December 2021, the majority of the 125 countries that make up the Convention on Certain Conventional Weapons were convinced of the need for restrictions. Russia and the US – heavy investors in the war industry – were the major opponents of regulations and questioned the need for a new LAWS treaty (Shead, 2021). China, paradoxically, wants to both continue the development and production of these weapons but ban their use. Other detractors included India, the UK and Israel. This resistance to banning or restricting LAWS is best explained by the fact that they offer new ways of expanding the already lucrative trade in weapons. The global market for AI in the military grew to \$8.58 billion in 2023 (Globe Newswire, 2023), and the US government, by far the world's largest arms dealer, doles out billions in military contracts not only to the world's five largest arms producers – Lockheed Martin, Boeing, Raytheon, Northrop Grumman, and General Dynamics – but also increasingly to many Silicon Valley start-ups (Schwarz, forthcoming).

The investment in LAWS has little to do with the 'good guys' defending us from the 'bad guys', but has everything to do with the latest expansion of a profitable war industry and the circumvention of international humanitarian law (the rules of war) at the expense of civilians around the globe. In 2019, the Trump administration profited from a massive sale of arms to Saudi Arabia and the United Arab Emirates. The \$8.1 billion 'emergency' arms deal circumvented Congress, which had tried to block it on the grounds that the Saudi-led coalition, which has been waging a war against the Houthis in Yemen since 2014, has been devastating (Robiou, 2019). Weddings, funerals and school buses have all been targeted, killing thousands of civilians while doing nothing to check the rise of the Houthis. The UN has verified that more than 10,200 children have been killed or injured in the ongoing conflict (United Nations, 2022a). The UK, Canada and France have also supplied the regime with weapons. Despite well-documented evidence, 21 members of the UN Human Rights Council, including Russia, Bahrain, and China, voted against the Dutch resolution to continue an independent investigation into war crimes in Yemen (Reuters, 2021).

Russia and Saudi Arabia have also been the major suppliers of arms to the Syrian government. A UN report documents that at least 306,887 civilians were killed between 1 March 2011 and 31 March 2021, while numerous independent international organizations have also documented the wide scope of war crimes committed in Syria (United Nations, 2022b). In January 2022, US President Biden, despite referring to President Abdel Fattah al-Sisi of Egypt as 'Trump's favorite dictator', approved a \$2.5 billion arms sale to the al-Sisi dictatorship even as America, along with independent organizations, criticized its record of abuse of human rights and suppression of democracy (Hansler, 2022). Germany and France have also supplied arms to Egypt.

Russia, following the same pattern as these other wars, likely believed itself immune to criticism when it attacked Ukraine in 2022, deliberately targeting civilian infrastructure and committing war crimes. The money raised in British financial markets – as Russian kleptocratic wealth flowed through London and oligarchs laundered money in shell companies and real estate – has funded Putin's war machine. Oliver Bullough, author of *Butler to the World: How Britain Became the Servant of Tycoons, Tax Dodgers, Kleptocrats and Criminals*, writes: 'Boris Johnson is congratulating himself on doing so much to help Ukraine, but Britain is like a doctor treating a patient's symptoms after causing the infection in the first place' (2022). So too, in 2021, EU countries, primarily Germany and France, sold weapons and ammunition worth €39 million ($42.3 million) to Russia despite the embargo that followed the Russian annexation of the Crimea (Guarascio, 2022). A critical part of its economy and as part of its 2023 arms deal, Iran has sold weapons to Russia and Russia has sold weapons to Iran, which has smuggled weapons to Hamas.

Following the 9/11 terrorist attacks, the war on terror in the Middle East, which has cost taxpayers trillions, made defence contractors, lobbyists and Pentagon officials enormously wealthy; defence stocks outperformed the stock market (Bilmes, 2021) while civilians bore the cost and the Houthis in Yemen, along with many other Arabs, were radicalized (Riedel, 2017: '2003: 'The tipping point' section). Millions around the globe protested the US invasion of Afghanistan and the occupation of Iraq (on the false premise that Iraq was developing weapons of mass destruction) to no avail. Aerial bombing and Predator and Reaper drone strikes killed thousands of civilians and destroyed much of the infrastructure. Although it is always difficult to get exact figures, one report that considers the public health costs of war in Iraq has estimated that there were '460,000 excess deaths from March 2003 to mid-2011' (Hagopian et al, 2013: 'Editor's summary' section). The destruction in war zones has also often not acknowledged long-term consequences with 'an estimated 3.6–3.8 million indirect deaths in post-9/11 war zones' where 'women and children suffer the brunt of these ongoing impacts' (Savell, 2023: 2).

Coming full circle, Israel cited America's response to 9/11 as its rationale for invading Gaza, following the murderous attacks and hostage taking of mostly Israeli citizens by Hamas, a designated terrorist group in the West, on 7 October 2023 (Mansoor, 2023). Just as 19 hijackers armed with box cutters murdered thousands of Americans, another of the most advanced militaries in the world, thanks to billions in American military aid, could not prevent the Hamas attacks, though the Israel Defense Forces (IDF) has been able to turn Gaza into a 'graveyard for children' (Wintour, 2023). Although the IDF claims that its AI-based system Habsora ('Gospel' in Hebrew) has reduced civilian deaths by generating more accurate targets, which has long been the promise of drones and LAWS, the facts tell a different story: one of maximum destruction and death in record time (Brumfiel, 2023). With the escalation of war rhetoric, as Schwarz has pointed out, 'the goalposts are shifting toward an acceptance of fully autonomous lethal machine decisions' and with this shift is an emphasis on speed, regardless of error, over ethics (Schwarz, forthcoming). The push for LAWS as part of the booming trillion-dollar arms trade, which blurs the distinction between the military and for-profit industry, has also trumped ethical discussions.

Just as with the COVID-19 pandemic, where AI proponents positioned their technology as critical to health care and were quick to hype its machine-learning algorithms as lifesaving with next to no results, so too are they now arguing that AI is critical to national security (Heaven, 2021). With wars in Ukraine and the Middle East, Silicon Valley is hounding at the door and hoping to cash in not only in America but also in Europe:

> Companies that sell military AI make expansive claims for what their technology can do. They say it can help with everything from the mundane to the lethal, from screening résumés to processing data from satellites or recognizing patterns in data to help soldiers make quicker decisions on the battlefield. Image recognition software can help with identifying targets. (Heikkilä, 2022: 'Why AI' section)

Eric Schmidt, the former CEO of Google and the founding member of the government's Defense Innovation Board, is pushing the American government to 'disrupt' the military and adopt AI technology: '"Let's imagine we're going to build a better war-fighting system," Schmidt said "We would just create a tech company"' (quoted in Mok, 2023).

Give the existence of unregulated global financial markets, the international arms trade, the influence of Big Tech and the failure to uphold even the semblance of the international community's obligation to protect civilians and infrastructures and prosecute war crimes, it is difficult to distinguish between the 'good guys' and the 'bad guys'. Now that war has broken out in Europe and democracies around the world are under threat, will the lucrative weapons industry be better regulated and will the rhetorical outrage in the West about the violation of human rights translate into conscionable action that stops propping up dictators and tyrants with arms? If the US, by far the largest arms dealer in the world, continues to invest in LAWS and spreads ever more deadly weapons around the world, we can likely expect more civilian deaths if the manufacturing and exporting of weapons follows the same logic as that of the gun industry

(Suchman, 2024). With the highest level of civilian-owned guns in the world, the US also has the most mass shootings and gun homicides (Mil, 2022). Both lucrative industries fight regulation, are backed by powerful lobbyists and use the same 'good guy' myth to distract from the carnage (Beauchamp, 2022).

2. The artificial intelligence industry and the military-industrial complex

In 2016 a prominent group in the AI industry penned an 'open letter' opposing the development of LAWS as if these systems could be hived off from this industry that has long been part of the military-industrial complex. It reads as follows: 'Starting a military AI arms race is a bad idea and should be prevented by a ban on offensive autonomous weapons beyond meaningful human control.' The letter is signed by many who have made their fortunes in the field of AI technology, including Elon Musk of SpaceX and Tesla; Steve Wozniak of Apple Inc.; Demis Hassabis of Google's DeepMind; Jack Dorsey of Twitter; Yann LeCun of Facebook AI Research; and Geoffrey Hinton of Google and the University of Toronto. The letter can be found on the website of the Future of Life Institute, an organization that was started with funding from Elon Musk and founded by Max Tegmark (Future of Life Institute, 2016).

Hinton, one of the signatories, moved to Canada from the Massachusetts Institute of Technology (MIT) in 1987, he explained, so that he could stop taking money from the US military, the major funder of most AI research in America, though he recognized the military would still use his research (Smith, 2017). One of the main architects of the artificial neural network that powers modern AI, he sold his company to Google for $44 million in 2012 and started to work half-time for them in 2013; he resigned in April 2023, citing his concerns about the technology and expressing regrets about his life's work (Heaven, 2023). Although the letter concludes 'we believe that AI has great potential to benefit humanity in many ways, and that the goal of the field should be to do so', Hinton has also claimed: '"I don't know any scientist who wouldn't explore an idea because it might have bad consequences"' (quoted in Barss, 2012). Working with the same belief as many others in the AI community that humans are machines (Onstad, 2018), Hinton helped develop image and voice recognition systems, which have been, among other uses, central to surveillance systems, LAWS and the rapid spread of disinformation. As Peter Galison has noted, once the human is reduced to an information machine, not only is privacy eroded, but we also risk 'the fabric of the democratic civil society that has taken so many years to construct and so many lives to defend' (Najafi and Galison, 2003).

Stuart Russell, a computer scientist at the University of California, Berkeley, founder of the university's AI lab, and also one of the signatories of the letter, has also expressed regrets about his contributions to the development of LAWS: '"The A.I. community, myself included, we were sort of asleep at the wheel for a long time and we weren't really thinking about the ways it could be misused"' (quoted in Kessel, 2019). Though wary of the misuse of AI and advocating for a ban on LAWS, Russell may find himself in the same quandary as Hinton, as those funding and using the research are in conflict with his own goals: in 2022 he accepted a $125 million fund for AI research from Schmidt, one of the loudest voices calling for more funding of military AI (Knight, 2023). In contrast to Russell, others have refused to accept the Schmidt award (Shead, 2022).

The open letter protesting the development of LAWS also sidesteps the long involvement of the AI industry in the war industry. The American military created Silicon Valley, pouring money into the region during the Cold War and again in the 1980s with Ronald Reagan's Strategic Defense Initiative and the Defense Advanced Research Project Agency (DARPA)'s Strategic Computing Initiative. As Margaret O'Mara (2019: 260) writes:

> the future-tense California had never been without the government's invisible hand. … defense remained the big-government engine hidden under the hood of the Valley's shiny new entrepreneurial sports car, flying largely under the radar screen of the saturation media coverage of hackers and capitalists. Contracts for missiles and lasers and interceptors didn't get much airtime in the many studies considering the race to build 'the next Silicon Valley'.

LAWS are not 'autonomous', but are extensions of the history of the AI industry and its drive for automation. Norbert Wiener, one of the founders of cybernetics and a central figure in the development of AI, in his 1940 wartime research on automatic anti-aircraft guns, collapsed the pilot with his plane. His defence system used an early computer to predict the statistical probability of the flight path of a plane based on the previous ten seconds of the pilot's course: 'the pilot was so merged with machinery that his human/nonhuman status was blurred' (Najafi and Galison, 2003). Wiener later wrote 'as objects of scientific enquiry, humans do not differ from machines'(Rosenblueth and Wiener, 1950, 326); what began with the reduction of the enemy to an information machine had now expanded to include all humans. The investment in AI systems, like Habsora, following on from Wiener's war research, has allowed for the rapid escalation of the number of 'targets', from 50 a year to a 100 a day in Gaza, based on machine-generated probabilities (Baggiarini, 2023). The mounting civilian death toll of these wars, where even children are caught up as data in mechanized and automated systems and collapsed with buildings and rocket launchers, skirt the Geneva Conventions and their Additional Protocols that are meant to protect noncombatants.

Hinton and Russell join the many scientists before them who have regretted their complicity with the war industry (Heffernan, 2015: 68) including Wiener. In the wake of the Second World War, he protested against the militarization of science, fearing we would become 'the slaves of our own technical improvement'. In a 1945 letter to Karl T. Compton, he wrote of an 'acute attack of conscience' over his war work and wanted to give up science 'and find some way of living on [his] farm in the country'. In a 1947 letter to the *Atlantic Monthly*, which was published as 'A scientist rebels', he wrote: 'If therefore I do not desire to participate in the bombing or poisoning of defenseless peoples – and I most certainly do not – I must take a serious responsibility as to those to whom I disclose my scientific ideas.' He refused further funding from the military and, recognizing that technology is not neutral, called for more responsible science:

> The policy of the government itself during and after the war, say in the bombing of Hiroshima and Nagasaki, has made it clear that to provide scientific information is not a necessarily innocent act, and may entail the gravest consequences. One therefore cannot escape reconsidering the established custom of the scientist to give information to every person who may inquire of him. The interchange of ideas which is one of the great traditions of science must of course receive certain limitations when the scientist becomes an arbiter of life and death. (Wiener, 1947: 46)

The entangled history of the military and AI industry, the collapse of humans and machines and the anthropomorphic language attributed to them, an arms trade that profits at the expense of global citizens and the regrets of scientists who have become part of an ever-more deadly war machine situate LAWS in a network that calls into question their 'autonomy'. Literal readings of fiction, which have animated the technology, have further contributed to the short circuiting of this wider context.

3. The 'Terminator conundrum' versus what fiction can teach us

In discussions of LAWS and the future of war, the Terminator is the go-to reference for those both for and against their use. Air Force General Paul Selva, Vice Chairman of the Joint Chiefs of Staff, said in a 2016 presentation: '"We're not talking about cruise missiles or mines. But robotic systems to do lethal harm … a Terminator without a conscience"' (quoted in Anuradha, 2017). Debating the 'Terminator conundrum', the US Deputy of Defense Bob

Work, one of the main drivers behind this technology, asked: '"If our competitors go to the Terminators … and it turns out the Terminators are able to make decisions faster, even if they're bad, how would we respond?"' (quoted in Scharre, 2018: 125). Steve Olsen, the deputy branch head of the Navy's mine warfare office, commented: '"The last thing we want to see is the whole "Terminator going crazy" [scenario], so we're working very hard to take the salient steps to protect ourselves and others"' (quoted in Larter, 2019). In 2017, Rich Haridy reported on the announcement by the Kalashnikov Group, the Russian arms manufacturer, that it had developed a range of combat robots that are fully automated and that use AI to identify targets. He concluded that the '"Terminator conundrum" may have been an amusing thought experiment for the last few years, but the science fiction is quickly becoming science fact. Are we ready to give machines the authority to make life or death decisions?' (Haridy, 2017).

With the 2019 release of another film in the franchise, *Terminator: Dark Fate*, the BBC reported that AI researchers objected to the use of the Terminator analogy, arguing that it functions as no more than clickbait that misleads the public about AI and causes misplaced fears about out-of-control technology. Yoshua Bengio, often referred to as one of the 'godfathers of AI', objects to the *Terminator* films for several reasons: '"They paint a picture which is really not coherent with the current understanding of how AI systems are built today and in the foreseeable future … We are very far from super-intelligent AI systems"' (quoted in Shead, 2019). Russell also objects to this use of the Terminator analogy, as, he argues, it is unrealistic. He participated in the making of the short film *Slaughterbots*, which he thinks offers a more realistic portrayal of how the technology works. Paul Scharre, author of *Army of None: Autonomous Weapons and the Future of War,* agrees that the Terminator analogy is not helpful in discussions of AI weapons. In an interview with Lucas Perry, he said:

> the Terminator is like the first thing that comes up because it's such a common pop culture reference. It's right there in people's minds. So I think go ahead and for the listeners, imagine that humanoid robot in the Terminator, and then just throw that away, because that's not what we're talking about. (Perry, 2020)

However, whether embraced or rejected as an analogy for LAWS, the problem in both cases is the invocation of the Terminator that narrowly focuses on it as a technology. In rendering a metaphor literal – a weapon springing from the pages of fiction – the ethical force of these films that depict the techno-military-industrial complex as a dehumanising war machine is brushed aside. Watts and Bode (2023), in their discussion of the *Terminator* franchise and AI narratives, include 'fiction and non-fiction' in shaping the sociotechnical imaginary, while acknowledging that AI in fiction works as a metaphor and noting that Cameron said that the films were never about machines destroying the human race, but about the loss of our humanity. The Terminator is not a literal weapon, but a character that, along with a pantheon of fictional others, from talking lions to aliens to witches to angels to vampires, navigates the fluctuating borders about what it means to be human.

Reading fiction as continuous with reality accepts a reality that fiction, by announcing itself as fiction, challenges. The imaginative worlds of fiction, unlike science, make no claim to being factual. Exploiting the conventions of both horror and science fiction, Cameron's three *Terminator* films involve time travel and other impossible scenarios. When fiction is read literally or policy uses fiction as if it were continuous with current technology or when fiction is dismissed as unrealistic, we lose the enormous ethical potential of fiction to defamiliarize or unfix reality, to expose its operating ideologies and to imagine alternative ways of living on the planet. Read as fiction rather than literally, Cameron's films expose the myths propagated by the AI industry and the dangers of the techno-military-industrial complex, where billions of dollars flow between governments and industries that feed on war.

3.1 Myth 1: AI as the future of war

The Terminator (as do the other two Cameron films) underscores one possible future: a dystopic LA where machines have risen from the ashes of nuclear war and are ploughing over barren fields littered with human skulls. A Terminator from 2029 is sent back to 1984 to kill the unborn son of Sarah Connor, who will lead the resistance movement against the machines. One of the resistance fighters, Kyle Reese, who ends up fathering Connor's child, has returned to protect her and redirect the course of events. Reese relays a message to Sarah from her future son: 'There is no fate.' The Los Angeles of 1984 shares many of the features of the 2029 world – alleys full of homeless people, streets full of trash, highways jam-packed with machines, and stores bursting with weapons, where the Terminator can easily buy up all the weaponry it needs to randomly kill anyone who crosses its path. The government does nothing to alleviate this misery, but instead funnels money into tech companies, weapons, police and security, which are ubiquitous in the film.

Terminator 2: Judgment Day (1991) also foregrounds the investments in the military-tech industry, at the expense of its citizens, with the SAC-NORAD (an acronym for Strategic Air Command-North American Aerospace Defense Command) funding of Cyberdyne Systems, the largest supplier of military computer systems. When Sarah awakens from a nightmare in which a playground full of children are obliterated by a nuclear bomb, she carves the words 'no fate' into the picnic table just before she heads off to kill Miles Dyson, the director of special projects for Cyberdyne and the engineer responsible for the neural-net processor that becomes Skynet, the AI system designed to defend the US.

In *Terminator: Dark Fate* (2019), the cycle continues, as the film opens in a world of displaced migrant populations, refugees and poverty while billions are spent on corporate military technology, cages, surveillance cameras and border patrols, feeding global inequalities instead of addressing them. Cameron's *Terminator* films do not foretell the future, but defamiliarize the present in order to expose the operating ideology of governments that invest in lucrative technologies connected to a war industry rather than investing in a better world for citizens around the world. The constant reminder in the films that there is 'no fate' resists the corporate branding of AI as the future of war.

The Terminator is a 1980s response to the market for techno-science, the DARPA-backed growth of the computer industry, the Cold War and the threat of nuclear catastrophe. The year 1984, which saw the first film's release, was also the year Apple's famous commercial that referenced George Orwell was released: 'On January 24th, Apple Computer will introduce Macintosh. And you'll see why 1984 won't be like 1984.' While the ad positioned its rival IBM as 'big brother', this was a distraction and clever marketing as, in fact, both corporations are products of a taxpayer-funded government and a military invested in surveillance technology. The year 1984 was also when the doomsday clock, which weighs the threats to humanity from unchecked scientific and technological inventions, moved up to three minutes before midnight. In 2022, global military spending reached an all-time high. The US, responsible for almost 40% of global expenditure in this field, invests 'more than 10 times as much on its military' as it does 'on education' (Gelling, 2023).

An investor in several tech companies, including the start-up military contractor Rebellion Defense, Schmidt seems undeterred by conflicts of interest as he tries to reinvigorate a sluggish tech sector by lobbying for more government funding of Silicon Valley while serving on two government advisory boards and promoting AI as the future of war (Knight, 2023). Never broaching the question of whether we need a 'future of war' and comparing '"[AI-powered] autonomy and decentralized, distributed systems"' to nuclear weapons, Schmidt situates this technology in the catastrophic legacy of the two world wars, when corporations, scientists and engineers all invested in chemical weapons, machine guns, gulags, gas ovens, atomic bombs and death camps (quoted in Knight, 2023). Rather than opening up the future to social progress, the military-backed AI industry traps us in a bleak legacy.

Octavia Butler, who is often credited with predicting the future in *Parable of the Sower* and *Parable of the Talents*, novels that were published in the 1990s but that recount America's slide into fascism, rejected the idea of fiction as prediction: 'All I did was look around at the problems we're neglecting now and give them about 30 years to grow into full-fledged disasters.' She pointed out that 'writing novels about the future doesn't give me any special ability to foretell the future. But it does encourage me to use our past and present behaviours as guides to the kind of world we seem to be creating' (2000).

3.2 Myth 2: STEM is the answer to all problems

Cyberdyne begins as a tech manufacturing corporation that accepts major contracts from the American military. The corporation then develops Skynet, which uses AI to replace commercial and military pilots and to control other military systems, including nuclear missiles. The *Terminator* films critique the unholy alliance of scientific research, corporate profit and the military that escalates the threat of global disaster. AI practitioners who insist that the *Terminator* films are misleading fail to acknowledge the dangers of the techno-military-industrial complex that the films expose.

'In labs at the University of Cambridge, Facebook and Amazon, researchers fear *Terminator: Dark Fate* could mislead the public on the actual dangers of artificial intelligence' (Shead, 2019). This article, which involved consulting a prominent group of computer scientists, mostly men with close ties to the AI industry, who are often called on by the media to comment on the state of AI, fails to acknowledge the problem with conflating government-funded universities and corporations. The former's mandate is to carry out independent research that aims to benefit society, while the goal of the latter is to amass profit and answer to shareholders. Neil Lawrence, one of those interviewed in the article, like many in the field, moves between the academic and corporate worlds with ease – from Microsoft to the University of Sheffield to director of machine learning at Amazon in Cambridge to his current home at the University of Cambridge as the first DeepMind professor of machine learning.

Science, technology, engineering and mathematics (STEM) as the key to national prosperity and the solution to all societal problems was marketed by the American government at the beginning of the 21st century and was swiftly adopted by other countries around the world. However, instead of national prosperity, the US has experienced a substantial increase in wealth inequality, a massive concentration of wealth and power in Big Tech corporations and a decline in public infrastructure. The cycle of the government funding industry and the industry mandate to make money bypasses public interests. In 2012, Schmidt, then Chairman of Google, despite the corporation heavily benefiting from tax dollars, declared he was '"proud"' of tax dodging – '"It's called capitalism"', he said (quoted in Kavoussi, 2012). More recently, the now former chairman of Google gifted the Yale Jackson Institute for Global Affairs $15.8 million to establish the Schmidt Program of AI, Emerging Technologies, and National Power, facilitating the link between Silicon Valley and the national security community. Schmidt, one of the wealthiest men in the world (Sandler, 2022), is one example of 'the disastrous rise of misplaced power' that Dwight Eisenhower warned against in 1961, where military spending on the AI industry bypasses democracy and enriches individuals, who then amass enough power and wealth to direct research programmes in universities and government policies that in turn feed the AI industry and a tech elite.

3.3 Myth 3: Machines are intelligent

In sequel after sequel, the Terminators return in human form as ever more ruthless machines, embodying the very logic of the techno-military-industrial complex. The computer scientists Bengio and Russell, both of whom oppose the use of the Terminator metaphor, refer to Skynet, the global software system that controls the Terminator and triggers nuclear war, as 'super-intelligent' (Shead, 2019). However, relying on motion tracking, search modes and facial recognition to inflict their blunt violence, the machines in the films are never represented as intelligent. As Reese, who grew up in the wake of the war, describes them: 'The H-Ks [hunter-killers] use infrared so you still have to watch out. But they're not too bright' (*The Terminator*, 1984).

Cathy O'Neil, a data scientist with a PhD in mathematics and author of *Weapons of Math Destruction*, points out that, despite the hype, what goes under the name of AI is a model, a necessarily simplified version of a complex reality: 'opinions embedded in mathematics' (2017: 21). They can be useful tools, but as they strip away context and subtly, replicate bias and 'hammer complexity into simplicity', they can inflict a great deal of harm when corporations and governments use them to manage populations (2017: 208). While the Google head, Sundar Pichai, pushes for 'AI everywhere' (Helft, 2016), the technology, which reduces humans to a set of data points in its system, could well be described in the same terms as Reese describes the Terminator: 'It can't be bargained with. It can't be reasoned with. It doesn't feel pity or remorse or fear' (1984). Automation bias, where humans view machine-generated results as objective and trust automated systems even in the face of contradictory evidence, magnifies the harm.

Despite the promise of LAWS to be 'more precise' in its target and to reduce harm to those not involved, Israel's automated system has, as discussed earlier, produced a record number of civilian deaths in a record amount of time. Valuing efficiency and speed over intelligent engagement and diplomacy, the AI weapons industry wants to automate war, but it is automated stupidity, not super-intelligence that is the real threat or, as Patrick

Crogan puts it, 'a pathway toward an increasingly stupid, global geopolitical military engagement in the world' (2019, 106).

3.4 Myth 4: Artificial intelligence is human-like

In *Terminator 2,* when Miles Dyson, the scientist behind the neural-net processor, learns of his responsibility for three billion deaths, the 'genius' behind Cyberdyne protests: 'You're judging me on things that I haven't even done yet. How were we supposed to know?' Sarah, who fights the machines that want to prevent her from giving birth and then target her child, responds:

> How are you supposed to know? Fucking men like you built the hydrogen bomb. Men like you thought it up. You think you're so creative. You don't know what it's like to really create something; to create a life; to feel it growing inside you. All you know how to create is death and destruction. (1991)

Turing, one of the 'fathers' of AI (notably, there are no mothers), in his 1950 article 'Computing machinery and intelligence', replaces the question of thinking with imitation as he fantasizes about birthing a 'child machine'. Mimicry and deception are at the core of the imitation game; in the first version of the Turing test, the man lies in order to convince the judge he is a woman and in the second version, the computer needs to deceive the judge as it tries to prove it is a man. Turing's theoretical model, where 'intelligence' is stripped of emotion, reflection, judgement, thought and ethics, has spun off into AI generators that can clone voices and images, producing deepfakes and chatbots that sow division.

The *Terminator* films were quick to expose the dangers of Turing's creation of deceptive machines that launched the AI industry. In the first film, the Terminator uses voice technology to disguise itself as Sarah's worried mother in order to track her down and kill her, while shapeshifting machines that can take on any disguise, object or human, are key to the death and destruction in *Terminator 2*. The AI system Skynet is what, in one possible future, gives way to nuclear war where all humans become targets. Referring to the global tensions caused by AI technology that spreads disinformation and fuels war, Cameron remarked in a BBC interview: '"All Skynet would have to do is just deepfake a bunch of people, pit them against each other, stir up a lot of foment, and just run this giant deepfake on humanity"', adding '"It would actually look a lot like what's going on right now"' (quoted in Kan, 2022).

The AI industry, funded by the military and following in Turing and Wiener's footsteps, reinvents the human as machine-like and then markets machines as human-like, capable of autonomy, creativity and intelligence.

These anthropological trappings distract from the role LAWS play in the larger network of a lucrative dehumanizing military-tech industry.

4. Technology, war and *R.U.R.*

AI as a lucrative technology and AI as a killing machine have long been intertwined in the cultural imaginary. Long before *The Terminator*, the connection between profitable automation and war was dramatized in Karel Čapek's *R.U.R.: Rossum's Universal Robots* (1920). Published shortly after the futility and carnage of the First World War, the play ends with a robot uprising and the death of all humans except one. Čapek's generation had grown up with the promise that modern science and technology would give way to a more humane, peaceful and civilized world. But that promise soured with the invention of aerial bombings, poisonous gas, machine guns, and tanks that made the First World War one of the deadliest wars in history. Witnessing the bloodiest and most senseless of wars, Čapek exposed the faulty logic that linked technological acceleration to social prosperity. If the Old Rossum wanted to prove God absurd by producing an artificial man, the Young Rossum is all about industry, efficiency and profit. Positing that reducing humans to machines as part of a dehumanizing war industry, Čapek wrote about his play:

> Young Rossum is the modern scientist, untroubled by metaphysical ideas; scientific experiment to him is the road to industrial production. He is not concerned to prove but to manufacture. To create a Homunculus is a medieval idea; to bring it in line with the present century this creation must be undertaken on the principle of mass-production. Immediately we are in the grip of industrialism; this terrible machinery must not stop, for if it does it would destroy the life of thousands. It must on the contrary, go on faster and faster, although it destroys in the process thousands and thousands of other existences. Those who think to master the industry are themselves mastered by it; Robots must be produced although they are a war industry, or rather *because* they are a war industry. (Čapek, 1923, 79)

All the technological fantasies that propel the AI industry today are already there in this 1920s play. Young Rossum, an engineer, harnesses science to industry; thinking of his 'billions', he designs human-like robots that are faster, cheaper and more productive than humans. The central director of Rossum's Universal Robots, much like the current heads of AI corporations, promises a world where robots will do all the labour and produce an abundance of goods, putting an end to poverty and freeing up humans to be creative. In producing the 'ideal' worker, he designs a robot, echoing the decades of dehumanising techno-capitalism, as ‘an ever cheaper commodity the more commodities he creates’, that feels no love, no joy, no empathy and no pain (Marx, 2010: 270). The term ‘robot’ from *robota*, the Czech word for servitude, first appears in this play. Despite the persistent reading of the play as ‘centred upon robots’, Čapek insisted that ‘I was much more interested in men than in Robots’ (1923).

The company head’s utopic dream that human-like machines will spell the end of work and herald an age of convenience, comfort and wealth concludes with a small group of men on an island amassing an enormous amount of money. Just as the AI industry is dominated by ‘fathers’, there are no women on the island that manufactures robots. When Miss Glory arrives, Fabry, the engineer, tells her: ‘It’s great progress to give birth by machine. It’s faster and more convenient’ and that nature cannot keep up ‘with the modern rate of work’ (Čapek, 2004: 18). As the corporation grows rich and inequality balloons, robots are then produced and sold as soldiers to quell uprisings and kill off rebellious populations around the world. When the robots unite and decide to exterminate all humans, they arrive on the island to finish off the corporate heads. Busman, the director of marketing, continues counting the company’s millions, refusing to be distracted from profit, even in the end times. While Alquist insists that not only was it a crime to teach robots to fight, but also a crime to make robots in the first place, Domin responds, with the familiar arrogance of the male tech ‘genius’ and despite facing the last days of civilization and the death of millions, that ‘it was a great thing’ (Čapek, 2004: 53).

Although Fabry tells Miss Glory ‘that there is nothing farther from being human than a Robot’, she finds the robots and the directors indistinguishable, mistaking the organic robot Sulla for a human and the company men for robots. The difference only manifests when Sulla expresses no horror at the idea of being dissected. Alquist is the only human that survives, and he is commanded and then begged by the robots to manufacture more robots, but the formula has been destroyed and, so it seems, the robots, unable to reproduce, will also meet their demise. However, in the final scene, when Alquist threatens to dissect two robots, echoing the earlier scene with Sulla but in reverse, he bears witness to their capacity for feeling – they display love, laughter, tears and self-sacrifice – and his despair turns to joy as he recognizes that ‘life will not perish’ (Čapek, 2004: 84).

The robot in fiction is not a literal machine, but the place to negotiate what it is to be human. Standing at the edge of the vast graveyard of humanity, Hallemeirer (the head of the institute for robot psychology and education) laments: ‘It was a great thing to be a human being. It was something tremendous’ (Čapek, 2004: 68). If the fantasy of imagining the human as an efficient machine ends with the near-extinction of all life, ‘life’ itself manifests as feeling, love and empathy, none of which can be programmed or manufactured or harnessed for profit.

Conclusion

When fictional robots/cyborgs are conflated with real technology and science fiction is read as continuous with science fact, technology is animated, anthropomorphized, and discussed as autonomous, creative and intelligent. Not only does this conflation distract from the role that the AI industry and LAWS play in the global arms trade, but fiction is made to serve as tech propaganda while the literary imagination is marginalized.

As Silicon Valley hopes for record windfalls and venture capitalists are urging Congress to set aside at least $1 billion for their military start-ups, which, as Senator Elizabeth Warren has noted, "'makes President Eisenhower's warning about the military-industrial complex seem quaint"' (quoted in Lipton, 2023), *R.U.R.* and the *Terminator* films help to unpack the ideological underpinnings of a technology that reduces humans to machines for profit, which in turn feeds a dehumanizing war industry. LAWS need to be understood not in terms of debates about their supposed autonomy, but in a wider historical context of automation, AI and the lucrative war industry that has produced ever more deadly weapons in the name of defence while doing little to protect civilians and public infrastructure around the world.

Filmography

The Terminator (1984) Directed by James Cameron. US: Cinema '84, Euro Film Funding, Hemdale, Pacific Western Productions.

Terminator 2: Judgement Day (1991) Directed by James Cameron. US: Carolco Pictures, Pacific Western, Lightstorm Entertainment, Le Studio Canal+.

Terminator: Dark Fate (2019) Directed by Tim Miller. US: Paramount Pictures, Skydance Media, Twentieth Century Fox.

References

Anuradha, G. (2017) 'What is Terminator conundrum? "Killer robots" in military raise ethical concerns', *International Business Times*, 9 February. Available from: https://www.ibtimes.com/what-terminator-conundrum-killer-robots-military-raise-ethical-concerns-2585223 [Accessed 12 July 2022].

Army Technology (2023) 'Zala KYB strike drone, Russia', *Army Technology*, 23 January. Available from: https://www.army-technology.com/projects/zala-kyb-strike-drone-russia/ [Accessed 13 January 2024].

Bächle, T.C. and Bareis, J. (2022) '"Autonomous weapons" as a geopolitical signifier in a national power play: analysing AI imaginaries in Chinese and US military policies', *European Journal of Futures Research* 10(20). https://doi.org/10.1186/s40309-022-00202-w

Baggiarini, B. (2023) 'Israel's AI can produce 100 bombing targets a day in Gaza. Is this the future of war?', *The Conversation*, 8 December. Available from: https://theconversation.com/israels-ai-can-produce-100-bombing-targets-a-day-in-gaza-is-this-the-future-of-war-219302 [Accessed 14 January 2024].

Barss, P. (2012) 'Toronto researcher wins prize for work on artificial brain', *Toronto Star*, 2 April. Available from: http://www.thestar.com/news/insight/2012/04/02/toronto_researcher_wins_prize_for_work_on_artificial_brain.html [Accessed 10 January 2015].

Beauchamp, Z. (2022) 'A child can't be a "good guy with a gun"', *Vox*, 25 May. Available from: https://www.vox.com/2022/5/25/23140519/uvalde-school-shooting-nra-texas-good-guy-gun [Accessed 12 January 2024].

Bilmes, L. (2021) 'Where did the $5tn spent on Afghanistan and Iraq go? Here's where', *The Guardian*, 11 September. Available from: https://www.theguardian.com/commentisfree/2021/sep/11/us-afghanistan-iraq-defense-spending [Accessed 12 January 2024].

Brumfiel, G. (2023) 'Israel is using an AI system to find targets in Gaza. Experts say it's just the start', *OPB News*, 14 December. Available from: https://www.opb.org/article/2023/12/14/israel-ai-warfare-gaza-gospel/ [Accessed 12 January 2024].

Butler, O. (2000) 'A few rules for predicting the future', *Essence*, May. Available from: https://sevengoodthings.com/a-few-rules-for-predicting-the-future-octavia-e-butler/ [Accessed 6 July 2023].

Bullough, O. (2022) 'Russia can only afford its war in Ukraine because Britain helped raise the cash', *The Guardian*, 8 April. Available from: https://www.theguardian.com/commentisfree/2022/apr/08/britain-war-in-ukraine-britain-putin-regime [Accessed 12 July 2022].

Čapek, K. (1923) 'The meaning of R.U.R.', *The Saturday Review*, 21 July, p 79. Available from: https://archive.org/details/sim_saturday-review-uk_the-saturday-review_1923-07-21_136_3534/mode/2up [Accessed 13 January 2024].

Čapek, K. (2004) *R.U.R. Rossum's Universal Robots*, London: Penguin Classics.

Cave, S. and Dihal, K. (2019) 'Hopes and fears for intelligent machines in fiction and reality', *Nature Machine Intelligence*, 1(2): 74–78. https://doi.org/10.1038/s42256-019-0020-9

Crogan, P. (2019) 'Visions of swarming robots: artificial intelligence and stupidity in the military-industrial projection of the future of warfare', in T. Heffernan (ed) *Cyborg Futures: Cross-disciplinary Perspectives on Artificial Intelligence and Robotics*, Cham: Palgrave Macmillan, pp 89–112.

Etzioni, A. (2018) *Happiness Is the Wrong Metric: A Liberal Communitarian Response to Populism*, Cham: Springer.

Future of Life Institute (2016) 'Autonomous weapons open letter: AI & robotics researchers', *Future of Life Institute*, 9 February. Available from: https://futureoflife.org/open-letter/open-letter-autonomous-weapons-ai-robotics/ [Accessed 28 April 2023].

Gelling, P. (2023) 'The US spent $877 billion on military in 2022, a new report says – more than 10 times federal education spending', *Business Insider*, 25 April. Available from: https://www.businessinsider.com/us-defense-military-budget-compared-federal-education-spending-2023-4?op=1 [Accessed 17 November 2023].

Globe Newswire (2023) 'Artificial intelligence in military global market report 2023', 27 March. Available from: https://www.globenewswire.com/news-release/2023/03/27/2635174/0/en/Artificial-Intelligence-in-Military-Global-Market-Report-2023.html [Accessed 28 April 2023].

Guarascio, F. (2022) 'EU closes loophole allowing multimillion-euro arms sales to Russia', Reuters, 14 April. Available from: https://www.reuters.com/world/europe/eu-closes-loophole-allowing-multimillion-euro-arms-sales-russia-2022-04-14/ [Accessed 28 April 2023].

Hagopian, A. et al (2013) 'Mortality in Iraq associated with the 2003–2011 war and occupation: findings from a national cluster sample survey by the University Collaborative Iraq Mortality Study', *PLOS Medicine*, 10(10): e1001533. https://doi.org/10.1371/journal.pmed.1001533

Hansler, J. (2022) 'Biden administration authorizes $2.5 billion in arms sales to Egypt despite human rights concerns', *CNN*, 25 January. Available from: www.cnn.com/2022/01/25/politics/us-arms-sales-egypt/index.html [Accessed 28 April 2023].

Haridy, R. (2017) 'Kalashnikov's new autonomous weapons and the "Terminator conundrum"', *New Atlas*, 21 July. Available from: https://newatlas.com/kalashnikov-ai-weapon-terminator-conundrum/50576/ [Accessed 12 July 2022].

Heaven, W.D. (2021) 'Hundreds of AI tools have been built to catch Covid. None of them helped', *MIT Technology Review*, 30 July. Available from: https://www.technologyreview.com/2021/07/30/1030329/machine-learning-ai-failed-covid-hospital-diagnosis-pandemic/ [Accessed 12 July 2022].

Heaven, W.D. (2023) 'Deep learning pioneer Geoffrey Hinton has quit Google', *MIT Technology Review*, 1 May. Available from: https://www.technologyreview.com/2023/05/01/1072478/deep-learning-pioneer-geoffrey-hinton-quits-google/ [Accessed 2 May 2023].

Heffernan, T. (2015) 'The post-apocalyptic imaginary: science, fiction, and the death drive', *Africa English Studies*, 58(2): 66–79.

Heikkilä, M. (2022) 'Why business is booming for military AI startups', *MIT Technology Review*, 7 July. Available from: https://www.technologyreview.com/2022/07/07/1055526/why-business-is-booming-for-military-ai-startups/ [Accessed 12 July 2022].

Helft, M. (2016) 'Inside Sundar Pichai's plan to put AI everywhere', *Forbes*, 18 May. Available from: https://www.forbes.com/sites/miguelhelft/2016/05/18/inside-sundar-pichais-plan-to-put-ai-everywhere/?sh=37e131924a2e [Accessed 4 December 2024].

Kan, M. (2022) 'James Cameron: Skynet would destroy humanity with deepfakes, not nukes', *PC Mag*, 24 January. Available from: https://www.pcmag.com/news/james-cameron-skynet-would-destroy-humanity-with-deepfakes-not-nukes [Accessed 12 July 2022].

Kavoussi, B. (2012) 'Google chairman Eric Schmidt defends tax dodge: "it's called capitalism"', *Huffington Post*, 13 December. Available from: https://www.huffpost.com/entry/google-tax-dodge_n_2292077 [Accessed 28 April 2023].

Kessel, J.M. (2019) 'Killer robots aren't regulated. Yet', *New York Times*, 13 December. Available from: https://www.nytimes.com/2019/12/13/technology/autonomous-weapons-video.html) [Accessed 28 April 2023].

Knight, W. (2022) 'Russia's killer drone in Ukraine raises fears about AI in warfare', *Wired*, 17 March. Available from: https://www.wired.com/story/ai-drones-russia-ukraine/ [Accessed 12 July 2022].

Knight, W. (2023) 'Eric Schmidt is building the perfect AI war-fighting machine', *Wired*, 13 February. Available from: https://www.wired.com/story/eric-schmidt-is-building-the-perfect-ai-war-fighting-machine/ [Accessed 28 April 2023].

Larter, D.B. (2019) 'The US Navy says it's doing its best to avoid a "Terminator" scenario in quest for autonomous weapons', *Defense News*, 12 September. Available from: https://www.defensenews.com/digital-show-dailies/dsei/2019/09/12/the-us-navy-says-its-doing-its-best-to-avoid-a-terminator-scenario-in-its-quest-for-autonomous-weapons/ [Accessed 12 July 2022].

Lipton, E. (2023). 'New spin on a revolving door: Pentagon officials turned venture capitalists', *New York Times*, 30 December. Available from: https://www.nytimes.com/2023/12/30/us/politics/pentagon-venture-capitalists.html [Accessed 14 January 2024].

Mansoor, P. (2023) 'Opinion: 9/11 offers awful lessons for what could happen with a Gaza ground invasion', *Los Angeles Times*, 16 October. Available from: https://www.latimes.com/opinion/story/2023-10-16/israel-hamas-gaza-september-11th-response [Accessed 13 January 2024].

Marx, K. (2010) 'Economic and philosophic manuscripts of 1844' (M. Milligan, trans.), in *Marx and Engels Collected Works* (Digital edition, Vol. 3, pp 229–349), London: Lawrence & Wishart.

Mil, D. (2022) 'Domestic gun ownership around the world', *GraphicNews*, 25 May. Available from: https://www.graphicnews.com/en/pages/42747/firearms-civilian-gun-ownership-by-country.

Mok, A. (2023) '"AI's impact on warfare will be as big as nuclear weapons", says ex-Google CEO', *Business Insider*, 14 February. Available from: https://www.businessinsider.com/ai-impact-war-nuclear-weapons-eric-schmidt-google-ex-ceo-2023-2?op=1.

Najafi, S. and Galison, P. (2003) 'The ontology of the enemy: an interview with Peter Galison', *Cabinet*, 12. Available from: https://www.cabinetmagazine.org/issues/12/najafi_galison.php [Accessed 12 January 2023].

O'Mara, Margaret (ed) (2019) *The Code: Silicon Valley and the Remaking of America*, New York: Penguin.

O'Neil, C. (ed) (2017) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, New York: Broadway Books.

Onstad, K. (2018). 'Mr. Robot', *Toronto Life*, 29 January. Available from: https://torontolife.com/life/ai-superstars-google-facebook-apple-studied-guy/ [Accessed 14 January 2024].

Perry, L. (2020) 'On lethal autonomous weapons with Paul Scharre', *AI Alignment*, 16 March. Available from: https://futureoflife.org/2020/03/16/on-lethal-autonomous-weapons-with-paul-scharre/ [Accessed 13 July 2022].

Reuters (2021) '"We have failed Yemen": UN human rights council ends war crime probe', *The Guardian*, 7 October. Available from: https://www.theguardian.com/world/2021/oct/07/un-human-rights-council-votes-to-end-yemen-war-crimes-investigation [Accessed 28 April 2023].

Riedel, B. (2017) 'Who are the Houthis, and why are we at war with them?', *Brookings*, 18 December. Available from: https://www.brookings.edu/articles/who-are-the-houthis-and-why-are-we-at-war-with-them/ [Accessed 24 January 2024].

Robiou, M. (2019) 'What you need to know about Trump's $8 billion Saudi arms deal', *PBS Frontline*, 14 July. Available from: https://www.pbs.org/wgbh/frontline/article/saudi-arabia-arms-deal-trump-what-to-know/ [Accessed 28 April 2023].

Rosenblueth, A. and Wiener, N. (1950). 'Purposeful and non-purposeful behavior', *Philosophy of Science*, 17: 318–326.

Rosenberg, M. and Markoff, J. (2016) 'The Pentagon's "Terminator conundrum": robots that could kill on their own', *New York Times*, 25 October. Available from: https://www.nytimes.com/2016/10/26/us/pentagon-artificial-intelligence-terminator.html [Accessed 28 April 2023].

Sandler, R. (2022) 'Here are the richest tech billionaires 2022', *Forbes*, 5 April. Available from: https://www.forbes.com/sites/rachelsandler/2022/04/05/here-are-the-richest-tech-billionaires-2022/?sh=5d926ec15e37 [Accessed 12 July 2022].

Savell, S (2023). 'How death outlives war: the reverberating impact of the post-9/11 wars on human health', *Costs of War: Watson Institute, Brown University*, 15 May. Available from: https://watson.brown.edu/costsofwar/files/cow/imce/papers/2023/Indirect%20Deaths.pdf [Accessed 13 January 2024].

Scharre, P. (2018) 'A million mistakes a second', *Foreign Policy*, 12 September. Available from: https://foreignpolicy.com/2018/09/12/a-million-mistakes-a-second-future-of-war/ [Accessed 12 July 2022].

Schwarz, E. (forthcoming) 'The hacker way: moral decision logics with lethal autonomous weapons systems', in H. Glaser and P. Wong (eds) *Governing the Future: Digitalization, Artificial Design, Dataism*, Boca Raton: CRC Press.

Shead, S. (2019) 'Why Terminator: Dark Fate is sending a shudder through AI labs', *BBC News*, 25 October. Available from: https://www.bbc.com/news/technology-50166357 [Accessed 13 July 2022].

Shead, S. (2021) 'UN talks to ban "slaughterbots" collapsed – here's why that matters', *CNBC*, 22 December. Available from: https://www.cnbc.com/2021/12/22/un-talks-to-ban-slaughterbots-collapsed-heres-why-that-matters.html [Accessed 28 April 2023].

Shead, S. (2022) 'Eric Schmidt plans to give AI researchers $125 million to tackle "hard problems"', *CNBC*, 16 February. Available from: https://www.cnbc.com/2022/02/16/eric-schmidt-plans-to-give-ai-researchers-125-million-.html [Accessed 13 July 2022].

Smith, C.S. (2017) 'The man who helped turn Toronto into a high-tech hotbed', *New York Times*, 23 June. Available from: https://www.nytimes.com/2017/06/23/world/canada/the-man-who-helped-turn-toronto-into-a-high-tech-hotbed.html [Accessed 28 April 2023].

Suchman, L. (2019) 'Demystifying the intelligent machine', in T. Heffernan (ed) *Cyborg Futures: Cross-disciplinary Perspectives on Artificial Intelligence and Robotics*, Cham: Palgrave Macmillan, pp 35–61.

Suchman, L. (2024) 'The algorithmically accelerated killing machine', 21 January. Available from: https://robotfutures.wordpress.com/2024/01/21/the-algorithmically-accelerated-killing-machine/ [Accessed 24 January 2024].

Turing, A. (1950) 'Computing machinery and intelligence', *Mind*, 59(236): 433–460. Available from: https://doi.org/10.1093/mind/LIX.236.433 [Accessed 13 July 2022].

United Nations (2022a) 'New action plan to strengthen the protection of children affected by armed conflict in Yemen signed with the Houthis', 22 April. Available from: https://childrenandarmedconflict.un.org/2022/04/new-action-plan-to-strengthen-the-protection-of-children-affected-by-armed-conflict-in-yemen-signed-with-the-houthis/ [Accessed 28 April 2023].

United Nations (2022b) 'Syria: decade of brutal war left nearly 307,000 civilians dead', 28 June 2022. Available from: https://news.un.org/en/story/2022/06/1121532 [Accessed 28 April 2023].

Watts, T.F. and Bode, I. (2023) 'Machine guardians: The Terminator, AI narratives and US regulatory discourse on lethal autonomous weapons systems', *Cooperation and Conflict*. https://doi.org/10.1177/00108367231198155

Wiener, N. (1947) 'A scientist rebels', *Atlantic Monthly*, January, p 47. Available from: https://www.theatlantic.com/magazine/archive/1947/01/a-scientist-rebels/656040/ [Accessed 8 July 2023].

Wintour, P, (2023) 'Gaza becoming "a graveyard for children", says UN Secretary General', *The Guardian*, 6 November. Available from: https://www.theguardian.com/world/2023/nov/06/gaza-becoming-a-graveyard-for-children-says-un-secretary-general [Accessed 11 November 2023].

From the Reel to the Real: Narratives of Weaponized Artificial Intelligence Technologies in India

Ingvild Bode and Shimona Mohan

Introduction

Narratives of artificial intelligence technologies (AIT)[1] feature in many fictional works, such as novels, movies, TV series and video games. Narratives typically represent events in a sequential and often causal way, and suggest specific ways of making sense of the world (Onega Jaén and García Landa, 1999: 3; Wibben, 2011: 59; Bode, 2015: 47). Due to their function as human sense-making tools (Suganami, 1997; Koschorke, 2018) that make the increasing integration of AIT in various aspects of social life legible, narratives have come to be the focus of cross-cultural, interdisciplinary research in the humanities and social sciences (Cave et al, 2020a; Cave and Dihal, 2023). Cave and Dihal (2019) have, for example, compiled eight of the most prevalent AI narratives in the popular culture of the Anglophone West[2] and have discovered that these narratives typically alternate between utopian and hopeful or dystopian and fearful depictions. Narratives of weaponized AIT, specifically, have been less prominently studied – potentially because the entire study of AI narratives is a comparatively recent phenomenon. However, they deserve more consideration since the political and social implications of weaponized AIT represent an extreme

[1] We recognize the umbrella term 'AI' as being vague, imprecise and politically contentious (Holland, 2023; Tucker, 2023). In this chapter, we use a broad definition of AI as the attempt 'to create machines or things that can do more than what is programmed into them' (Gebru, 2023).

[2] We recognize the problematic connotations of the term 'Anglophone West' (see Mac Sweeney 2023), but use this descriptor as it appears in the original work by Cave and Dihal.

form of possible harm, including physical injury or death. The military application of these technologies is certainly not science fiction, as seen by rising state investments into AIT in the military domain (Bode and Huelss, 2022; Garcia, 2023; Scharre, 2023).

This chapter contributes to this literature by investigating narratives about weaponized AIT in India based on original survey data (conducted in January 2023).[3] India is an under-researched case in scholarship about weaponized AIT and autonomous weapons systems (AWS)[4] that remains focused on a small selection of countries, chiefly the US, China and, to a more limited extent, Russia (for example, Jensen et al, 2020; Kania, 2020; Nadibaidze, 2022; Bendett, 2023; Qiao-Franco and Bode, 2023). India represents an interesting case to examine for two reasons: first, Indian decision makers have begun to invest more significant financial resources into developing and acquiring weapons integrating autonomous or AI technologies in targeting (Boulanin and Verbruggen, 2017). Simultaneously, Indian delegations to the Group of Governmental Experts (GGE), the prime international forum where debate about AWS happens, have opposed the negotiation of a legally binding instrument. This departs from the stance expressed by the Non-Aligned Movement,[5] which has supported a move towards legally binding regulation. Rather, India's more techno-optimist position aligns with those expressed by US and European developers of weapon systems integrating autonomous and AI technologies in stating that 'autonomy in critical functions of a weapon system would impart more precision and accuracy and would avoid human errors' (CCW Group of Governmental Experts on LAWS, 2021).

This chapter connects India's policy-making stance with an analysis of narratives about weaponized AI, building on scholarship that suggests how AI narratives may influence the global governance of AIT (Cave and Dihal, 2019; Hudson et al, 2023; Watts and Bode 2023). We examine survey responses provided by Indian participants via three research questions: (1) to what extent, if any, are Indian respondents familiar with weaponized AI narratives derived from the AI narratives categorized in Cave

[3] This effort is part of a larger survey covering French, Japanese and US publics conducted as part of the ERC AutoNorms project. This chapter draws on data analysis first included in Bode et al (2024).

[4] In the following, we refer to weaponized AI and AWS interchangeably. However, we acknowledge that AWS may, but do not necessarily, integrate AI technologies.

[5] The Non-Aligned Movement (NAM) originated as part of decolonization in the 1950s to represent the interests of Global South countries that had recently gained their independence. In the specific context of the Cold War, NAM countries chose not to enter any military alliances with one of the two big powers of the time. The NAM has retained its significance as a concerted forum for countries of the Global South.

and Dihal's research (2019) of fictional products associated with the Anglophone West?; (2) are there alternative narratives of weaponized AIT particular to India that go beyond existing narrative categories? By researching how the Indian cultural context and embedding foster different narratives of weaponized AI, this undertaking helps to develop an understanding of how AIT are imagined globally (Bareis and Katzenbach, 2022; Cave and Dihal, 2023); and (3) to what extent do such weaponized AI narratives shape the Indian respondents' attitude towards regulating the weaponization of AIT?

The remainder of the chapter is structured as follows: first, we review literature on AI narratives to locate our contribution. As our analytical framework, we modify the eight AI narratives categorized by Cave and Dihal (2019) to concentrate explicitly on representations of and fears about weaponized AI. Second, we offer a brief introduction to India's national military AI ecosystem and its positioning in the international debate on AWS, especially in relation to discussion at the GGE. We find that while India has only recently begun to be active in its engagements around military AI, its activity within this domain is steadily accelerating. Third, we reflect on the survey's methodology. Fourth, we present, discuss and analyse our survey findings. Indian respondents displayed significant familiarity with weaponized AIT narratives associated with the Anglophone West. But they also shared distinctly Indian narratives of weaponized AIT that blend elements of Indian mythology and folklore with commonplace Western themes such as the *Frankenstein* motif. Such weaponized AIT narratives appeared to be influential for Indian respondents' attitudes towards regulating this technology, but our findings on this are not conclusive.

1. Narratives of (weaponized) artificial intelligence

The growing integration of AI technologies in our daily lives has been preceded by how artificially 'intelligent' systems are imagined in various works of fiction and nonfiction. Histories of science fiction have long documented European and American works most prominently (Mukherjee, 2020: 4). Mary Shelley's 1818 novel *Frankenstein* with its 'machines as monsters' motif is thus taken as one of the earliest expressions of this genre (Hermann, 2023: 322; Hudson et al, 2023: 197), while scholars also note that such narratives date back much longer to, for example, Homer's *Iliad* (Cave et al, 2020b). Work conducted by researchers of the AI Narratives project at the University of Cambridge's Leverhulme Centre for the Future of Intelligence has been pivotal in expanding this space. Again, initially, this scholarship primarily explored AI narratives in the Anglophone West, but, more recently, has also covered global narratives, including those originating in India (Mukherjee, 2023).

This connects to studies of Indian science fiction that describe both particular regional contexts (Mukherjee, 2020) and draw comparisons across the country's diverse languages and regions (Banerjee, 2020; Khan, 2021; Khilnani and Bhattacharjee, 2022; Kuhad, 2022). These recent works offer rich insights into Indian science fiction and are intended to counterbalance Eurocentric perspectives on the genre (Khan, 2021: xiii). With some notable exceptions (Sondhi, 2022; Mukherjee, 2023), scholarship rarely speaks in detail specifically to narratives of AI. Broadly, Indian science fiction is characterized by a distinct and particular blending of modernity, folk tales, postcolonial motifs and religious beliefs (Banerjee, 2020). Sondhi's examination of the Hindi-language movie *Anukul* (2017) underlines this. Set in a world where humans and humanoid robots coexist, the movie depicts the eponymous humanoid robot Anukul. Anukul is capable of learning ethical values associated with the 'Bhagavad Gita, one of the foremost religious texts in Hinduism', but also shows a pre-programmed murderous intent (Sondhi, 2022: 208).

In the past ten years, research on AI narratives has gained significant sway and attention across the social sciences and the humanities for at least two reasons. First, the extent to which some of the 'imagining' associated with AI technologies has arguably been 'realized' has increased the issue's salience. Second, scholars have become much more attentive to how narratives shape public discourse about emerging technologies, such as AI, by informing what kind of meanings and potential futures are associated with these technologies (Cave and Dihal, 2019: 74; Dillon and Schaffer-Goddard, 2023; Hudson et al, 2023). Scholars hold that such narratives can influence how AI technologies are developed, adapted, funded, understood and regulated.

Researchers of the AI narratives project have produced the most comprehensive catalogue of common AI narratives that can be found in more than 300 fictional and nonfictional works chiefly associated with the Anglophone West (Cave et al, 2019). As summarized in Table 1, these narratives can be categorized into representing 'hopeful' and 'fearful' visions of AI. In fact, such hopeful and fearful categorizations are flipsides of each other. The hopeful narrative of 'ease' therefore imagines AI technologies as providing labour for humans and offering relief from the burden of work. This is a pervasive form of imagining AIT that becomes visible, for example, in the depiction of various robot servants such as *The Jetsons'* Rosey (1962). Showing the other side of the picture, the 'obsolescence' narrative speaks to the unintended consequences of eliminating the need for human labour: humans losing their sense of purpose and responsibility for self-care. Here, *Wall-E*'s (2008) depiction of obese humans moving around in multi-entertainment sensory pods offers a useful example.

When it comes to narrative depictions of weaponized AIT, the fearful flipside tends to dominate. The *Terminator* franchise's dystopian depiction

Table 1: Major AI narratives identified in fiction and nonfiction of the Anglophone West

Description	AI narratives		Description
	Hopeful categorization	**Fearful categorization**	
AI significantly extends human life expectancy and quality of life.	Immortality	Inhumanity	AI enables certain aspects of human consciousness to separate from the body, but at the cost of the individual's basic humanity.
AI provides labour on command, relieving humans from the burdens of work.	Ease	Obsolescence	By eliminating the need for human labour, AI undermines an individual's sense of purpose, self-worth and responsibility for self-care.
AI fills the social need for friendship, compassion and/or romance.	Gratification	Alienation	AI leads people to reduce if not eliminate their social interaction with other human beings.
AI is used to preserve and/or extend certain utopian ways of life.	Dominance	Uprising	AI escapes human control and sets about humanity's destruction.

Source: Based on Cave et al (2019)

of a robot uprising, associated with powerful and widely circulated visuals such as the red-eyed T-800, in particular, has become a master narrative of weaponized AI in the Anglophone West (Carpenter, 2016: 53; Scharre, 2018: 264; Payne, 2021: 13; Watts and Bode, 2023).

We use this categorization of AI narratives as an analytical backdrop, but adapt it to focus on a subset of four narratives that we consider speaking most directly to weaponized AI. These are the narrative pairs dominance and uprising, as well as gratification and alienation. Table 2 summarizes these narratives, including the descriptions we shared with respondents in the survey. We modified descriptions offered by Cave, Coughlan and Dihal in their survey of the UK public[6] to highlight aspects particularly relevant for the specific context of AIT in weapon systems.

[6] We needed to modify the descriptions because the original versions did not focus on weaponized AI. Our modifications stayed true to the original themes as discussed in the AI narratives project's publications, but included some more detail. The original descriptions of the eight AI narratives used in their survey of the UK public are included in Table 1.

Table 2: Four narratives of weaponized AI used in the survey

Adapted description	AI narratives		Adapted description
	Hopeful categorization	**Fearful categorization**	
The development of weaponized and other forms of AI can provide humans with various forms of companionship including friendship, service and guardianship.	Gratification	Alienation	The development of weaponized and other forms of AI raises philosophical questions about what it means to be human, and whether being human can be learnt by machines.
The development of weaponized and other forms of AI can produce smarter and more capable weapon systems. These can be used to both attack and protect the humans which created them.	Dominance	Uprising	The development of weaponized and other forms of AI could produce a malevolent super-intelligence that is more powerful than humans and overthrows its creators.

Source: Adapted from Cave et al (2019)

Later on, we will provide additional information about the design of our questionnaire survey and our methodological choices.

2. India and military artificial intelligence technologies

While India is renowned as a massive information technology (IT) hub, its engagements with critical new and emerging technologies like AI are still limited in terms of scope and activity. Interestingly, Indian defence research and development (R&D) had already established the Centre for Artificial Intelligence and Robotics (CAIR) under the Defence Research & Development Organization (DRDO) in 1986, long before AIT as we now know it existed. However, the Centre produced very little activity around AI for defence purposes until India's uptick in AI-related advancements from around 2018 onwards.

NITI Aayog, which is the primary public policy think tank of the Government of India, came out with an exploratory and sector-agnostic National Strategy on AI in 2018, which sparked off activity around application-specific AI engagements for India as well (NITI Aayog, 2018). In the same year, India convened a multi-stakeholder task force around formulating a strategy for AI for national security, which resulted in the setting up of a Defence Artificial Intelligence Council (DAIC) under the Ministry of Defence (MoD), and a Defence AI Project Agency (DAIPA) under the Department for Defence Production (DDP) of the Ministry of Defence (MoD, 2022a).

While this institutionalization resulted in an increase in dedicated funding for defence AI products and indigenous development of military technology enhanced by AI, there has been little activity around a specific national policy or strategy for military AI.[7] In the meantime, the Ministry of Defence released a catalogue of 75 AI-based defence products and technologies which India launched in 2022 (MoD, 2022b), and each of the three services (that is, the Indian Army, Navy and Air Force) have undertaken fragmented actions around incorporating military AI within their own purviews.

India has been active in the Convention on Certain Conventional Weapons Group of Governmental Experts on Lethal Autonomous Weapons Systems (CCW GGE on LAWS) since its inception in 2017. India held the presidency of the Group for two consecutive years: in April and November 2017, and again in April and August 2018. The work done during these years resulted in the only substantive output from the GGE process, that is, the 11 Guiding Principles on LAWS (CCW Group of Governmental Experts on LAWS Report, 2019). India still regularly makes interventions at the Group's annual sessions, which have stressed the importance of the application of international humanitarian law on LAWS, but have increasingly tended to oppose a complete ban on them or their regulation, instead preferring lenient regulations (Human Rights Watch, 2020).

In 2023, another global process around LAWS and military AI more broadly was convened by the Netherlands, entitled the Responsible AI in the Military Domain (REAIM) Conference, where India was also in attendance. However, India did not end up signing the call of action that emerged from the conference, which about 75 per cent of the other countries that attended the conference did. The call included a statement by these countries to incorporate responsible AI into the development and deployment of their military AI ecosystems, and also to support this consideration when it is brought up in global discussions around LAWS like the CCW GGE (Government of the Netherlands, 2023).

---

[7] The Indian government does not have a documented geopolitical approach towards military applications of AI. The 2022 catalogue of AI-based defence products and technologies mentioned in the text is currently the most relevant military AI document that India has produced, and even within this, India's aspirations are melded and diluted with allusions to become a 'global AI hub', without specific mention of being a military AI power. The 2018 national AI strategy mentioned in the text, which is now largely considered to be outdated, is also sector-agnostic and has almost no mention of India's aspirations or engagements with AI for military, defence or security purposes, barring fleeting mentions of cybersecurity and individual data privacy and security.

Instead, at the time of writing (January 2025), India joined countries like the US, Russia, China and Israel in being more vocal about its tacit endorsement for the development of LAWS in the absence of stringent international regulations, hinting at the incumbent augmentation of its national military AI architecture. India's abstention from signing the call is also indicative of a larger trend of India prioritizing its national security considerations over international cooperation around tactical weapons. This is also evidenced by India's refusal to sign the Nuclear Non-Proliferation Treaty (NPT), which it believes is a discriminatory framework that 'does not constitute or contribute to the development of customary international law' (Ministry of External Affairs, 2021).

3. Methodology

We obtained our data from an online questionnaire survey conducted in India in January 2023. We will now proceed to describe the survey's content, the conduct of the survey via YouGov and the data limitations. Despite being anonymous, the survey asked several broad questions about gender, age and location to gauge the sample's diversity.

Ten questions were included in the survey, which was divided into three parts.[8] Part 1 asked about fundamental concepts and visual representations of weaponized AI,[9] Part 2 surveyed familiarity and salience with the four narratives about weaponized AI listed in Table 2 as well as other narrative

[8] The questionnaire was conducted in English.

[9] Part 1 included four questions. Q1: Artificial intelligence (AI) describes 'machines that can perform tasks that require human-level intelligence'. AI is developed through applications of mathematical logic, advanced statistics and/or forms of machine learning. AI can be used for many purposes (for example, smart assistants, self-driving cars and chatbots) and is considered a 'dual-use' technology because it has many commercial uses. Militaries across the world are investing in the use of various forms of weaponised AI to support military personnel and fight wars. Why do you think this is the case? This was an open-ended question. Q2: Which of the following images most closely corresponds to your understanding of weaponized AI? Respondents were shown five different images without captions (depicting a humanoid killer robot, autonomous drone, computer vision, human-machine interaction and animal-form robots). For each picture, respondents could choose between one of five options: closely matches, matches, does not match, does not match at all, don't know). Q3: Weaponized AI has been part of many (Western) television and film series. Please tick all the following television and film series which you have seen. Respondents could choose any of the following list or the option 'none of these'. Multiple answers were explicitly allowed. *2001 A Space Odyssey*, *The Avengers*, *Battlestar Galactica*, *Black Mirror*, *Bladerunner*, *Ex Machina*, *I Robot*, *The Matrix*, *Robocop*, *Star Trek*, *Star Wars*, *The Terminator*, *Transformers*, *Wall-E*, *Westworld*. Q4: Do you think the above television and film series portray weaponized AI in a universal way or do you think there is something particularly 'Western' or 'American' about them? Please explain your answer in a couple of sentences. This was an open-ended question.

themes,[10] and Part 3 asked about attitudes towards regulation.[11] Five multiple-choice items and five open-ended questions were included in the survey, which used a mixed-structure methodology. Our survey's open-ended questions, which differ from past, typically quantitative surveys of the public's attitude on AWS (for example, Horowitz, 2016; Ipsos, 2021; Rosendorf et al, 2022), were a crucial component for us. Through the inclusion of these, we intended to gain a deeper understanding of the reasons why Indian respondents may have a favourable or unfavourable opinion about the development of weaponized AI. Given the chapter's focus on AI narratives and regulatory attitudes, we only present empirical results associated with a subset of five questions from Parts 2 and 3 of the survey (Table 3).

We worked via the commercial survey provider YouGov to access a representative sample of the general public aged 18 and older residing in India. Following the usual YouGov panel process, respondents were rewarded with 'points' for their participation. We received a total of 510 respondents from the Indian public. The survey took an average of three to four minutes to complete.

A qualitative-interpretivist methodology guides how we handled survey data. Instead of statistically assessing this data, we seek to understand the responses as contextual forms of meaning making. In order to accomplish this, we used an inductive technique to code the responses to the open questions (Q8 and Q10) and iteratively debated and altered our coding themes, for instance, by going through specific cases and contrasting coding approaches. All the responses included in the next section are reproduced verbatim.

4. Summary of findings and analysis

Q6: Books, comics, films, television shows, video games and other fictional products tell various stories about how AI may be used in war. Which of the following types of stories are you familiar with?

Respondents could choose among one of three options (yes/no/unsure) for each of the four weaponized AI narratives as described in Table 2.

[10] The questions in Part 2 correspond to research questions 1 and 2: (1) to what extent, if any, is the Indian public familiar with weaponized AI narratives derived from the AI narratives categorized in Cave and Dihal's research (2019) of fictional products associated with the Anglophone West? (2) Are there alternative narratives of weaponized AIT particular to the Indian public that go beyond existing narrative categories?

[11] The questions asked in Part 3 correspond to research question 3: to what extent do such weaponized AI narratives shape the Indian public's attitude towards regulating the weaponization of AIT?

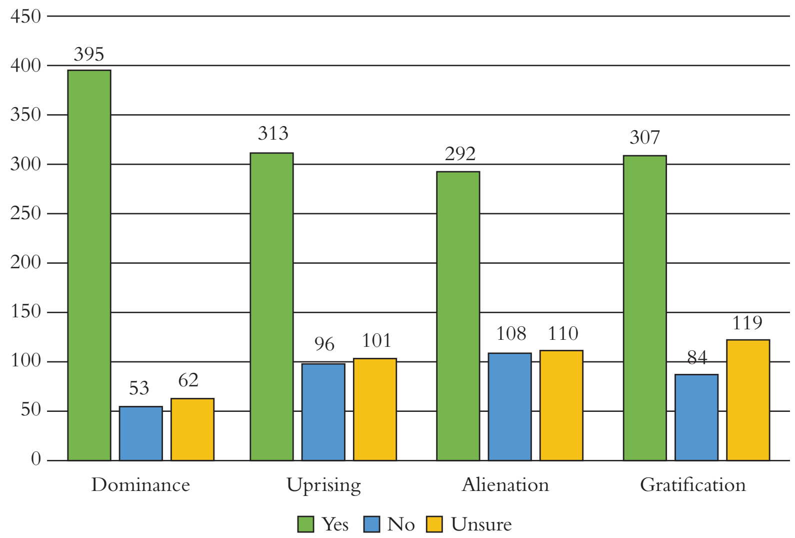
Table 3: Selected survey questions

Part 2: Narratives about weaponized AI	Q6. Books, comics, films, television shows, video games and other fictional products tell various stories about how AI may be used in war. Which of the following types of stories are you familiar with? *Respondents could choose among one of three options (yes/no/unsure) for each of the four weaponized AI narratives as described in Table 2.*
	Q7. Which of these stories about AI has influenced your thinking the most, if any? *Respondents could choose any of the four weaponized AI narratives or the option 'none of them'. The survey featured the descriptions outlined in Table 2.*
	Q8. Do you think that the above stories overlook any other important implications of AI, including the use of AI in war, which you have experienced in Indian popular media featuring these technologies? *This was an open-ended question where respondents could enter their own text.*
Part 3: Regulation	Q9. Have the portrayals of weaponized AI in popular culture influenced your thinking on whether such technologies should be regulated in real life? *Respondents could choose between the options yes/no/unsure.*
	Q10. Can you please explain your answer? In what ways, if any, have the portrayals of weaponized AI in popular culture influenced your thinking on whether such technologies should be regulated in real life? *This was an open-ended question where respondents could enter their own text.*

Generally, Indian respondents showed a high degree of familiarity with the four 'weaponized AI narratives' dominance, uprising, alienation and gratification (Figure 1). Most respondents were familiar with the dominance narrative, closely followed by the alienation and the uprising narratives. Overall, Indian respondents were least familiar with the gratification narrative. This may be because the gratification narrative is perhaps not as immediately associable with forms of weaponized AI, but may rather be chiefly connected to other forms of AI technologies – for example, voice/virtual assistants such as Samantha in the 2013 film *Her*.

It is also notable that Indian respondents rarely used the option 'unsure'. Given that the weaponized AI narratives we surveyed originate in works associated with the Anglophone West, this finding across English language-speaking Indian respondents is insightful. It potentially also tells us something about the extent to which these weaponized AI narratives travel throughout the Anglophone world outside of the West. At the same time, as responses to Q8 show, Indian respondents also put forward many additional fictional works, mostly movies, as sources of weaponized AI narratives. Therefore, the responses of the Indian survey participants may suggest a generalized

Figure 1: Familiarity with weaponized AI narratives



high degree of familiarity with some extremely well-known popular culture products and the weaponized AI narratives these products feature.

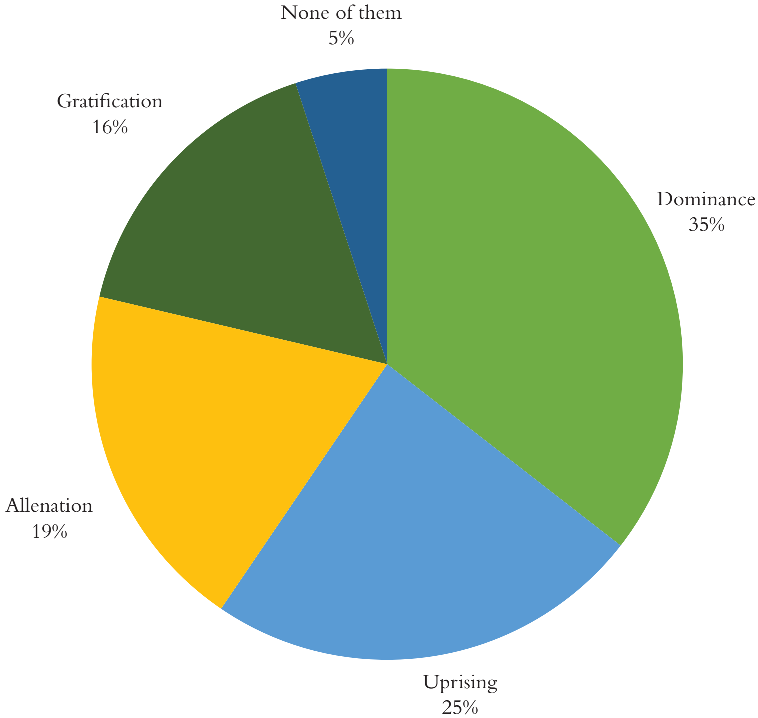
Q7. Which of these stories about AI has influenced your thinking the most, if any?

Indian respondents indicated that the dominance narrative influenced their thinking on weaponized AI the most, followed by uprising, and alienation with gratification being identified as least influential. These results reflect the high levels of familiarity respondents indicated regarding the four weaponized AI narratives. Among Indian respondents, the dominance narratives stand out very clearly, but overall levels of influence attributed to the other three narratives are high (Figure 2). In fact, only 5 per cent of respondents to the Indian survey noted that none of these narratives influenced their thinking.

Q8. Do you think that the above stories overlook any other important implications of AI, including the use of AI in war, which you have experienced in Indian popular media featuring these technologies?

Given that the four narratives whose familiarity we queried are based on the Anglophone West, we created Q8 to elicit responses addressing other perspectives and narratives concerning weaponizing AI that may be culturally distinctive. We organize responses to Q8 in two ways: first, Figure 3 summarizes responses received by Indian survey participants in numerical terms and, second,

Figure 2: Results

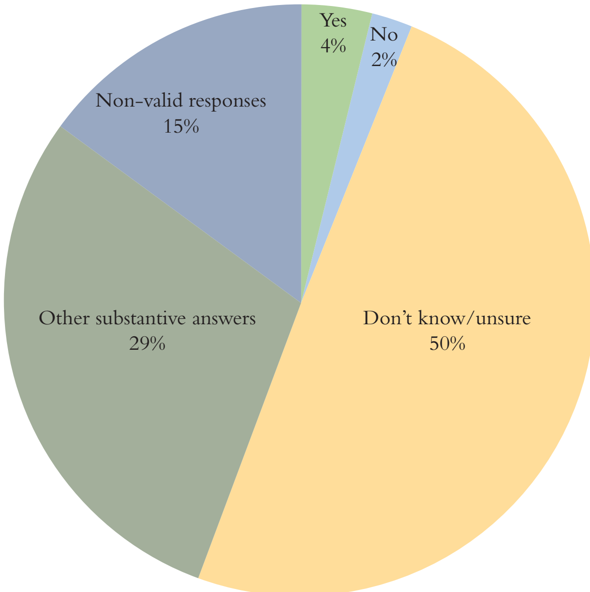


we then look at responses in the category 'substantive answers' (39 per cent) in more detail.

As visualized in Figure 3, the majority of Indian respondents (52 per cent – the number combines respondents simply stating 'yes', as well as other, substantive answers) think that the way in which the four narratives present weaponized AI overlooks important issues. A total of 18 per cent of respondents indicated that these four narratives did not overlook anything substantive, while 13 per cent of respondents expressed uncertainty.[12]
'Other substantive answers' is a category that collects together a range of responses of various lengths and speaking to various themes. We created an inductive coding pattern with ten major categories to give a structured

[12] We coded responses as invalid when respondents wrote down arbitrary words or pasted text from internet sources. Research on invalid survey responses highlights several reasons for this, including carelessness/inattentiveness, misrepresentation and linguistic incompetence (Johnson, 2005: 105). Research has further identified invalid responses as being due to a lack of motivation as a particular problem in commissioned surveys conducted by commercial companies (Curran, 2016: 4).

Figure 3: Overview of responses to open question



analysis of the responses in this category. Some of the codes describe narratives that the four narratives of weaponized AI offered in Q6 (Table 2) do not or only partially cover. Other codes highlight broader considerations of both military and civilian uses of AI. Furthermore, we chose to distinguish between assessments of AI as positive or negative rather than hopeful or fearful because the former captures a broader range of evaluations. To show the range of responses provided by Indian respondents, we used 12 coding categories. No one category is dominant among Indian respondents.

To structure the summary of our findings for Q8, we discuss three overarching themes that were present to significant degrees among responses provided by Indian respondents: (A) references to Indian popular culture products and references to India; (B) negative assessments of AI (including a discussion of the codes 'other AI narratives, 'malice, and 'malfunctions'); and (C) positive assessments of AI.

A. References to Indian popular culture products

More than a quarter of Indian respondents named specific popular culture products, primarily films, as providing alternative narratives about weaponized AI to those presented. Many respondents mentioned South Indian films and, in particular, Tollywood. Tollywood (distinct from the homonymous Tollywood film industry of the Bengali-speaking community in East India) is a film industry based in the South Indian states of Andhra Pradesh and Telangana which produces Telugu-language movies. However, the term is often used as a misnomer by people outside of the South Indian sphere to refer to all the South Indian cinematographic industries collectively, including their outputs in several other vernacular Indian languages like Tamil. South Indian movies are known for their higher propensity of generating fanciful, 'out-there' fare, with flashy computer graphics and science fiction or mythological fantasy themes in their plots.

The most frequently cited films were *Robot*, *2.0* and *Ra.One*. Because of how prominently these films figured in responses, we provide a brief plot overview of these films. *Robot*, also known as *Enthiran* in the original Tamil (2010), is set in present-day India. It revolves around Chitti, an android humanoid robot designed by scientist Dr Vaseegaran (in his likeness) for use by the Indian Armed Forces. Chitti initially fails to receive approval by the Artificial Intelligence Research and Development Committee (AIRD) due to insufficiencies in its 'neural schema' that manifest in the potential for homicidal behaviour. Vaseegaran subsequently modifies Chitti, allowing it to understand and display human emotions. As a result of these modifications, Chitti is approved for use by the Indian Army, but also falls in love with Vaseegaran's fiancée, Sanaa. Frustrated and humiliated by his invention, Vaseegaran dismantles it. At this point, rival scientist Bora rebuilds Chitti from its discarded pieces, installing a red chip into its neural schema that makes it malicious. Chitti 2.0 then goes into full 'uprising' mode – it abducts Sanaa, kills its creator Bora, makes replicas of itself to create an army, and fights a battle with government troops at the AIDR headquarters, leading to multiple casualties. As part of the battle, Chitti 2.0's army of robots engage in networked, swarm-like battle behaviour. Vassegaran manages to apprehend Chitti 2.0 and to remove the red chip. The film ends with the trial and acquittal of Vaseegaran, who instructs Chitti to deactivate itself. After apologising to its creator and to Sanaa, Chitti complies. Its dismantled body parts remain on confined display at an AI museum.

Ra.One (2011), a Hindi-language superhero film, is set in a future, virtual-reality world. Its plot revolves around game designer Shekhar Subramanium, who creates a motion sensor-based game around the virtual avatar, G.One (designed in Shekhar's likeness), and its antagonist, Ra.One (faceless), who is designed to be substantially more powerful. Ra.One powers include shape-shifting and he has access to self-learning AI. Shekhar's son Prateek begins playing the game under the identity Lucifer and soon becomes the only player to be able to keep up with Ra.One. Enraged by this, the self-aware Ra.One exits the game's virtual world and enters the real world, using novel technology introduced at the beginning of the film, with the purpose of killing Lucifer. Ra.One kills Shekhar and continues to pursue his family,

Table 4: Coding categories for 'other substantive answers' in alphabetical order

Code	Short description	Examples of responses
Ethics	Respondents reflect explicitly or implicitly about ethical implications of AI, using keywords such as 'emotions' or 'empathy'.	*We tend to forget humans are the ones who have created such magnificent resources to help us all. AI is the future but it wouldn't be possible without humans.*
Malfunctions	Respondents mention malfunctions, unreliability or distrust.	*Software glitches, issues with connections etc may impact its performance.*
Malice	Respondents note that other implications of AI are related to malicious intentions of actors.	*AI once developed with human-like attributes of feelings and emotions can become lethal or dangerous – if put in the wrong hands.*
Media critique	Respondents criticize the media for not providing sufficient information about or an unrealistic portrayal of AI.	*Indian popular media have chosen to ignore AI.*
Negative assessment of AI	Respondents demonstrate negative reflections on AI.	*Use of AI for destruction is not palatable for me.*
Other AI narratives as identified by Cave and Dihal	Respondents refer to alternative narratives as categorized by Cave and Dihal.	*Many have concerns about how advances in AI will affect what it means to be human, to be productive, and to exercise free will* [refers to alienation narrative, authors' addition].
Others	Respondents provided broader comments which are difficult to categorize.	*They are making UAV* [uncrewed aerial vehicles].
Positive assessment of AI	Respondents provide a positive assessment of the uses of AI in the civilian or the military domain.	*Yes, human harming will be reduced, no more losses.*
References to India	Respondents frame their answers with specific reference to India.	*Indians don't fear AI.*
Reference to Indian popular culture products	Respondents mention Indian popular culture products as sources of alternative AI narratives.	*Robot; 2.0; Ra-one; South Indian movies (Tollywood)*
Uncertainty	Respondents say they are unsure about implications of AI for various reasons – for example, they do not have enough information or they believe that this is a futuristic topic.	*AI is very much new to Indian culture so many people don't have knowledge about it.*
Value-neutral comments about AI	Respondents mention the uses of AI in other areas outside of the military, but do not evaluate those.	*AI can be used *to medical as well.*

who manage to summon G.One from the virtual world to defend them against Ra.One. Eventually, G.One succeeds in destroying Ra.One with the help of Prateek and takes himself back into the game's virtual world. The film ends with Prateek managing to return G.One to the real world and him taking Shekhar's place. Interestingly, the protagonist-antagonist nature of the characters is also depicted by their nomenclature – Ra.One is a homonym of *Raavan* (the main villain in the Indian epic, the *Ramayana*), whereas G.One is pronounced the same as *Jeevan* (the Hindi word for 'life').

2.0 (2018), a sequel to *Robot*, *2.0*, revisits Vaseegaran eight years later, where he has just created android humanoid assistant Nila. When the city's mobile phones begin to soar into the sky and subsequently regroup in a swarm-like fashion to kill people, Vaseegaran is urged to re-activate Chitti. In a more fantastical rather than science-fiction plot twist, the mobile phones turn out to be animated by the aura of Pakshi (a literal translation of 'bird' in Hindi), a recently deceased ornithologist, who criticized excessive mobile phone use, believing that high-frequency electromagnetic radiation from cell sites endangered bird life. Chitti, along with an army of robots, fights Pakshi's spirit in several encounters and eventually destroys it. The film ends with Chitti and Nila along with the army of robots continuing to serve the Indian government.

Robot and *2.0* were, respectively, the most expensive film productions of their time and, as their prominent appearance in our survey indicate, appear to have left an indelible mark on Indian public consciousness. These films echo some of the Anglophone West's weaponized AI narratives, such as uprising (Mahalakshmi and Rajendran, 2023). Both *Robot* and *Ra.One* revolve around a variation of the *Frankenstein* plot, associated with Mary Shelley's idea of a monster whose origins can be traced back to the use of scientific knowledge (Lakkad, 2018: 237): 'Victor Frankenstein assembles a body from various parts of fresh corpses and then endows it with life. He quickly rejects the new being, which disappears and becomes a threat to him and others' (Aldiss and Wingrove, 2001: 46–47). But the way in which *Robot* engages with the Frankensteinian theme differs from, for example, the well-known plot of *The Terminator*. Vaseegaran creates Chitti with the explicit purpose of military use and specifically to replace Indian soldiers, thereby seeking to alleviate the harmful effects of war in the form of traumatized soldiers and broken families (Lakkad, 2018: 241). The film contrasts Vaseegaran's patriotic vision of science with the malevolent vision of science purported by Bora and expressed by the red chip. In other words, the film presents a somewhat balanced vision of the development of AI as both the source of utopia and dystopia, but arguably more forcefully underlines fears associated with robotic technologies in warfare (Kaur, 2013; Lakkad, 2018).

Ra.One also features the Frankensteinian theme in the form of the eponymous invincible, shape-shifting protagonist reminiscent of the T-1000 in *Terminator 2: Judgement Day* (Watts and Bode, 2023). Comparable to *Robot*'s presentation of diverging visions of science, 'Ra.One and G.One are products of the same technology' (Vatsala and Raut, 2019: 231). But *Ra.One* blends these science-fiction themes with characteristics associated with Indian myth, such as the *Ramayana* and the *Mahabharata* (Vatsala and Raut, 2019: 227–9). Notably, a few Indian respondents to the survey explicitly mention the latter as presenting alternative Indian narratives about weaponized AI. The *Mahabharata* is one of the most prominent ancient Indian texts, and describes the story and the aftermath of the ambiguously fictionalized dynastic struggle called the Kurukshetra War. The text depicts a variety of war scenes, and ideological motifs from the *Mahabharata* are abundantly present throughout Indian history, political thought, literature, Hindu philosophy and public consciousness. Presumably, the survey respondents intended to refer to some of the mythical weaponry mentioned in the *Mahabharata*, such as the *Sudarshana Chakra* (the celestial discus) that could autonomously find its target and attack it, safely returning to its owner thereafter. While it functions in a similar manner to the kind of autonomous weapons we discuss in military AI conversations today, the *Chakra* draws its 'autonomous' capability from the divine energy of its wielder in the *Mahabharata*, Lord Krishna, who can then use it as an extension of his own being and command it at will. This juxtaposition between a frontier technology and mythical divine weaponry in the public perception of military AIT presents an interesting contrast between Indian and Western narratives of weaponized AI.

In sum, while popular films featuring weaponized AI named by Indian respondents share some of the themes present in AI narratives of the Anglophone West, there are also important differences, pointing to distinct imaginings of AI in India that blend elements of mythology, religion and modernity (Kaur, 2013). Some of these distinctions may also arise from the way in which weaponized AIT is presented in Indian and Western films, as well as the plot of the film itself. Indian films usually encompass a plethora of themes within the same production (action, drama, comedy, romance and the distinctive musical aspect), while Western films tend to be more genre-specific. The latter allows the viewers to focus exclusively on the plot and see weaponized AIT with the gravity of realism, while the former portrays them in a fictionalized format that affords them a veil resembling surrealism. Perhaps some of the Indian respondents' perceptions of AIT are coloured by this aggrandized and unrealistic version presented to them through the films they watch.

B. Negative assessments of artificial intelligence

Indian respondents expressed negative opinions of AI, which may be a sign of widespread anxiety: *'use of AI for destruction is not palatable for me'*; *'as the* *capabilities of using AI increase infinitely, there can be unimaginable consequences of using AI in war'*; *'it may not to be permitted to release all alone and may make destruction'*; or *'AI can't be use in war, it will erased human'*. Some of the worries expressed are based on broader AI narratives as categorized by Cave and Dihal, such as alienation (*'Many have concerns about how advances in AI will affect what it means to be human, to be productive and to exercise free will'),* obsolescence ('*Since India is a country with a lot of manpower, AI can actually create unemployment and other issues'*) or uprising (*'That could be devastating of AI learned how to enslave humans'; 'AI can be good until it starts thinking on its own which seriously lead to a grave situation'*). The last response can be read as an implicit reference to the *Robot* film's final scene, when Chitti states what it perceives to be the reason for being dismantled: 'I had begun to think' (Lakkad, 2018: 247). In fact, several respondents' negative assessments appear to echo malicious science themes depicted in *Robot*: *'No, it's dangerous to humans, after all it's machine, it works according to us until it runs perfectly, but many of us have bad inventor, if someone changed the way, machine will destroy humanity'* or *'AI once developed with human-like attributes of feelings and emotions can become lethal and dangerous – if put in the wrong hands'*.

However, Indian respondents appear to frequently take a 'balanced' stance in assessing AI, recognizing both positive and negative depictions of weaponized AI – for example, *'I think there are both risks and benefits available over these technologies'*; '*AI in war can be helpful and protective but there can be bad effects'*; or *'War destroys everything … But AI will be very helpful in war'*.

C. Positive assessments of artificial intelligence

Finally, several Indian respondents express fully positive opinions about the use of AI in both civilian and military scenarios: *'Its significant impact on human lives is resolving some of society's most pressing issues'*; or *'AI would be very popular whether in common people technology or in the modern weapons because it requires less manpower and is guided by the engineers with a lot of distance'*. Further, Indian respondents proved to be comparatively enthusiastic about the positive visions associated with AI in warfare: *'AI is very helpful to our Indian Army for protecting our people from the war'*; *'Human harming will be reduced, no more losses'*; or '*It is very useful to war and reduces the human job'*.

Such responses broadly correspond to those provided by Indian respondents in other, international surveys on AWS, for instance, the Ipsos surveys commissioned by the Campaign to Stop Killer Robots (2018 and 2020). In both cases, India is the only country where the majority of respondents support autonomous weapons, with 56 per cent in 2020 and 50 per cent in 2018 (Ipsos, 2019, 2021). Considering that India is sandwiched geopolitically between neighbouring countries with which it has political conflicts and border skirmishes are common news items, some of the respondents' perceptions around weaponized AIT being useful in war-like situations may stem from a securitized nationalistic viewpoint.

Q9. Have the portrayals of weaponized AI in popular culture influenced your thinking on whether such technologies should be regulated in real life?

Q10. Can you please explain your answer? In what ways, if any, have the portrayals of weaponized AI in popular culture influenced your thinking on whether such technologies should be regulated in real life?

More than half of the Indian respondents (57 per cent) affirmed that their attitude towards regulation has been influenced by portrayals of weaponized AI in popular culture, while 23 per cent responded 'no' and 20 per cent were unsure.

However, as responses to the open Q10 show, when specifically asked to explain their answer and reflect on the ways in which the portrayals of weaponized AI in popular culture have influenced their regulatory attitudes, Indian respondents are not as certain.[13] A total of 50 per cent of Indian respondents answered 'I don't know', while another 15 per cent gave nonvalid answers, which featured copied-and-pasted text from online publications and what appear to be AI-generated answers. This makes it difficult to draw firm conclusions regarding how Indian respondents relate depictions of weaponized AI in popular culture to their regulatory attitudes. However, similar to Q8, we inductively coded seven kinds of replies in the 'other substantive answers' category to demonstrate diversity in answers provided. Respondents illustrated, to some extent, ties between popular culture and their attitudes towards regulation with responses like *'The biggest movies and series invariably show the harmful side of weaponized AI violating the three laws of robotics [of Isaac Asimov]'* or *'It has shown the negative side, so it makes me think'*. Only one Indian respondent specifically opposed regulation, but the aggregate number of negative views of AI technologies is comparatively low. According to some Indian respondents, AI is *'a risky investment whether it is in monetary terms of intellectual and emotional ways'* or *'one day robots will take over us instead of helping us'*. Many respondents also emphasize the importance of regulation without overtly tying this attitude to popular culture, for example, simply noting that AI *'must be regulated'* or making more elaborate statements such as *'AI should be weaponized to save the people rather than kill and spread Chaos, there should be an international regulatory body for AI moderation'*. Many

[13] The authors want to express their gratitude to Anna Nadibaidze for her assistance in coding the survey data for Q10.

Table 5: Coding categories of 'other substantive answers'

Code	Description	Examples of responses
1. Popular culture shows regulatory needs	Respondents suggested that representations of weaponized AI in popular culture demonstrate regulatory needs in real life.	*I think it's unpredictable what will happen if we use AI in real life and I think so because as I have watched in movies.*
2. Regulation is needed, but no references to popular culture	Respondents identified regulatory needs but did not refer to popular culture.	*This technology needs to be regulated to prevent destructive uses of AI instead of being used for good reasons.*
3. Opposition to regulation, no regulation needed	Respondents demonstrate opposition to regulating AI technologies.	*AI and machinery should not be regulated in real life, because it may cause confusion.*
4. Negative assessments of AI	Respondents demonstrate opposition to regulating AI technologies.	*Weaponized AI should not be promoted in the real life. Definitely they transform as threat to the ecosystem. AI can be only implemented to monitor.*
5. Positive assessment of AI	Respondents provide a positive assessment of the uses of AI in the civilian or in the military domains.	*Using robots helps reduce loss of life – there are AIs that guide people through troubled times.*
6. Dismissal of popular culture	Respondents do not consider popular culture to be influential in their thinking about regulating AI.	*All this *look so imaginary that I cannot relate to all such stories though they are interesting.*
7. Others	No clear attitude towards popular culture or regulation could be identified.	*AI technology amuses me every single time.*

Indian respondents put forward positive assessments of AI, for example, believing that AI *'is good for the army'* and *'helps reduce loss of life'*, while weaponized AI *'is useful to protect higher officials'*.

5. Discussion of findings

We draw out six observations across the answers provided by Indian respondents in line with our three research questions.

First, the narratives of weaponized AI classified by Cave and Dihal in the Anglophone West also resonate among an Indian audience. The levels to which these narratives resonate vary, but they appear to be relevant reference points. In fact, Indian respondents also associated popular cultural goods from the

Anglophone West, such as the *Terminator* franchise, with the uprising narrative. This demonstrates the apparent, actual salience of such weaponized AI narratives outside of their original cultural contexts. Of course, parts of the Indian respondents, especially the tiny fraction of urban, English-educated middle and upper-class milieu, may also consider themselves as part of the Anglophone West, or at least more initiated into it compared to their Indian peers.

Second, some weaponized AI narratives appear to be more salient than others. Among the Indian public surveyed, attitudes towards weaponized AI were more closely connected to 'fearful' than 'hopeful' narratives.

Third, despite the overall salience of AI narratives associated with the Anglophone West, Indian respondents also put forward a range of alternative narratives of weaponized AI that are firmly associated with Indian popular culture and science fiction, in particular movies. A narrative analysis of movies such as *Robot* and *Ra.One* underlines the ways in which these blend narrative themes associated with the Anglophone West, such as the Frankenstein motif, with distinctly Indian themes, including mythologies and those motivated by the different imagining of AIT tied to the mixing of film genres.

Fourth, Indian respondents reached positive and negative as well as balanced assessments of AIT. In this, Indian respondents surveyed seem to differ from other global publics that demonstrate a predominance of negative assessments (Ipsos, 2021; Bode et al, 2024). At the same time, many Indian respondents still appear to have an unfavourable view of AIT. Responses to Q8 and Q10 tended to emphasize the concerns about AIT and the hazards they can be associated with, albeit to varying degrees. This is an intriguing finding because it indicates, for example, that cultural practices and attitudes towards robotics and AI technology in civilian settings do not necessarily translate into public understandings of appropriateness when it comes to robotic or AI-driven weapon systems.

Fifth, Indian respondents indicated that popular culture influenced their perception of the risks involved with weaponized AI, with some openly indicating that popular culture also influenced their regulatory attitudes. But the majority of Indian respondents supported regulation without mentioning popular culture as specifically influential for their attitudes. As a result, we cannot make solid judgements about the extent to which popular culture influences regulatory attitudes on AWS among the Indian public.

Finally, despite the many interesting answers provided, there appears to be somewhat of a lack of *substantive* knowledge and awareness of weaponized AI and associated narratives among Indian respondents. This is evidenced, in particular, by answers provided to the survey's open-ended questions. In Q10, for example, more than half of the Indian respondents chose the 'don't know' option. This indicates either a lack of information or a refusal to engage with the question. It may also indicate that the Indian respondents' general understanding of how AIT are already integrated into military systems remains limited.

Conclusion

Building on original data derived from a public opinion survey conducted in India in January 2023, we examined: (1) whether and to what extent narratives related to weaponized AI originating in the Anglophone West resonate in India; (2) whether Indian respondents identify specific, alternative narratives of weaponized AI as influential in their thinking; and (3) whether these narratives influence Indian respondents' views on regulating AI technologies.

Our findings imply that narratives identified in the popular culture of the Anglophone West have resonance in India. Some narratives, such as dominance and uprising, have a stronger resonance than others. Indian respondents also present several alternate narratives of weaponized AI, including many nuanced and positive assessments. At the same time, several of the cultural distinct narratives identified by Indian respondents share narrative themes already identified in English-language and Western popular culture. Moreover, the responses show that 'fearful' AI narratives are also quite prevalent in India, as is a lack of awareness about AIT, accompanied by some anxiety about what their use entails. Yet, overall, respondents from the Indian public appear quite confident in their knowledge about AIT, including those integrated into weapon systems. In this way, Indian respondents, at least to some degree, appear to share the techno-optimist perspective on weaponized AIT that is also purported by governmental elites.

Filmography

2.0 (2018) Directed by S. Shankar. India: Lyca Productions.

Enthiran (Robot) (2010) Directed by S. Shankar. India: Sun Pictures.

Her (2013) Directed by Spike Jonze. US: Annapurna Pictures.

Ra.One (2011) Directed by Kanika Dhillon and Niranjan Iyengar. India: Red Chillies Entertainment.

The Terminator (1984) Directed by James Cameron. US: Hemdale & Pacific Western Productions.

Terminator 2: Judgment Day (1991) Directed by James Cameron. US: Carolco Pictures, Pacific Western Productions, Lightstorm Entertainment & Le Studio Canal+.

Wall-E (2008) Directed by Andrew Stanton. US: Walt Disney Pictures & Pixar Animation Studios.

References

Aldiss, B.W. and Wingrove, D. (2001) *Trillion Year Spree: The History of Science Fiction*, London: Collins.

Banerjee, S. (2020) *Indian Science Fiction: Patterns, History and Hybridity*, Cardiff: University of Wales Press.

Bareis, J. and Katzenbach, C. (2022) 'Talking AI into being: the narratives and imaginaries of national AI strategies and their performative politics', *Science, Technology, & Human Values,* 47(5): 855–81. DOI: 10.1177/01622439211030007

Bendett, S. (2023) 'Roles and implications of AI in the Russian-Ukrainian conflict', *Russia Matters*, 20 July. Available from: https://www.russiamatters.org/analysis/roles-and-implications-ai-russian-ukrainian-conflict [Accessed 1 August 2023].

Bode, I. (2015) *Individual Agency and Policy Change at the United Nations: The People of the United Nations*, Abingdon: Routledge.

Bode, I. and Huelss, H. (2022) *Autonomous Weapon Systems and International Norms*, Montreal: McGill-Queen's University Press.

Bode, I., Huelss, H., Nadibaidze, A. and Watts, T. (2024) 'Cross-cultural narratives and imaginations of weaponised artificial intelligence: comparing France, India, Japan, and the United States', *Big Data & Society*, 11(4). https://doi.org/10.1177/20539517241303151

Boulanin, V. and Verbruggen, M. (2017) *Mapping the Development of Autonomy in Weapon Systems*, Stockholm: SIPRI Report.

Carpenter, C. (2016) 'Rethinking the political / -science- / fiction nexus: global policy making and the Campaign to Stop Killer Robots', *Perspectives on Politics*, 14(1): 53–69. https://doi.org/10.1017/S1537592715003229

Cave, S. and Dihal, K. (2019) 'Hopes and fears for intelligent machines in fiction and reality', *Nature Machine Intelligence*, 1(2): 74–78. https://doi.org/10.1038/s42256-019-0020-9

Cave, S. and Dihal, K. (2023) *Imagining AI: How the World Sees Intelligent Machines*, New York: Oxford University Press.

Cave, S., Coughlan, K. and Dihal, K. (2019) *Scary Robots: Examining Public Responses to AI*, Cambridge: Leverhulme Centre for the Future of Intelligence, Available from: http://lcfi.ac.uk/media/uploads/files/AIES-19_paper_200_Dihal_Cave_Coughlan_XLLBdft.pdf [Accessed 23 January 2024].

Cave, S., Dihal, K. and Dillon, S. (2020a) *AI Narratives: A History of Imaginative Thinking about Intelligent Machines*, Oxford: Oxford University Press.

Cave, S., Dihal, K. and Dillon, S. (2020b) 'Introduction: imagining AI', in S. Cave, K. Dihal and S. Dillon (eds) *AI Narratives: A History of Imaginative Thinking about Intelligent Machines*, Oxford: Oxford University Press, pp 1–24.

CCW Group of Governmental Experts (GGE) on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (LAWS) (2019). Available from: https://documents.unoda.org/wp-content/uploads/2020/09/CCW_GGE.1_2019_3_E.pdf [Accessed 23 January 2024].

CCW Group of Governmental Experts on LAWS (2021) 'Statement by India', *United Nations*. Available from: https://conf.unog.ch/digitalrecordings/?mrid=5EBAAABD-DD16-458E-A2C9-FEA476FCE99B [Accessed 8 August 2023].

Curran, P.G. (2016) 'Methods for the detection of carelessly invalid responses in survey data', *Journal of Experimental Social Psychology*, 66: 4–19. DOI: 10.1016/j.jesp.2015.07.006

Dillon, S. and Schaffer-Goddard, J. (2023) 'What AI researchers read: the role of literature in artificial intelligence research', *Interdisciplinary Science Reviews*, 48(1): 15–42. DOI: 10.1080/03080188.2022.2079214

Garcia, D. (2023) *Common Good Governance in the Age of Military Artificial Intelligence*, Oxford: Oxford University Press.

Gebru, T. (2023) 'Don't fall for the AI hype', *Tech Won't Save Us*, 19 January. Available from: https://techwontsave.us/episode/151_dont_fall_for_the_ai_hype_w_timnit_gebru [Accessed 19 June 2023].

Government of the Netherlands (2023), 'REAIM 2023 endorsing countries and territories'. Available from: https://www.government.nl/documents/publications/2023/02/16/reaim-2023-endorsing-countries [Accessed 23 January 2024]

Hermann, I. (2023) 'Artificial intelligence in fiction: between narratives and metaphors', *AI & Society*, 38(1): 319–329. DOI: 10.1007/s00146-021-01299-6

Holland, M.A. (2023) 'Recalibrating assumptions on AI', *Chatham House*, 12 April. Available from: https://www.chathamhouse.org/2023/04/recalibrating-assumptions-ai [Accessed 23 January 2024].

Horowitz, M.C. (2016) 'Public opinion and the politics of the killer robots debate', *Research & Politics*, 3(1): 1–8. DOI: 10.1177/2053168015627183

Hudson, A.D., Finn, E. and Wylie, R. (2023) 'What can science fiction tell us about the future of artificial intelligence policy?', *AI & Society*, 38(1): 197–211. DOI: 10.1007/s00146-021-01273-2

Human Rights Watch (2020) 'Stopping killer robots: country positions on banning fully autonomous weapons and retaining human control', 10 August. Available from: https://www.hrw.org/report/2020/08/10/stopping-killer-robots/country-positions-banning-fully-autonomous-weapons-and [Accessed 23 January 2024].

Ipsos (2019) 'Six in ten (61%) respondents across 26 countries oppose the use of lethal autonomous weapons systems', 23 January. Available from: https://www.ipsos.com/en-us/news-polls/human-rights-watch-six-in-ten-oppose-autonomous-weapons [Accessed 23 January 2024].

Ipsos (2021) 'Global survey highlights continued opposition to fully autonomous weapons', 2 February. Available from: https://www.ipsos.com/en-us/global-survey-highlights-continued-opposition-fully-autonomous-weapons [Accessed 1 February 2023].

Jensen, B.M., Whyte, C. and Cuomo, S. (2020) 'Algorithms at war: the promise, peril, and limits of artificial intelligence', *International Studies Review*, 22(3): 526–550. DOI: 10.1093/isr/viz025

Johnson, J.A. (2005) 'Ascertaining the validity of individual protocols from Web-based personality inventories', *Journal of Research in Personality*, 39(1): 103–129. DOI: 10.1016/j.jrp.2004.09.009

Kania, E. (2020) '"AI weapons" in China's military innovation', *Brookings Institution*, April. Available from: https://www.brookings.edu/wp-content/uploads/2020/04/FP_20200427_ai_weapons_kania_v2.pdf [Accessed 23 January 2024].

Kaur, R. (2013) 'The fictions of science and cinema in India', in K.M. Gokulsing (ed) *Routledge Handbook of Indian Cinemas*, Abingdon: Routledge, pp 282–296.

Khan, S.A. (2021) *Star Warriors of the Modern Raj: Materiality, Mythology and Technology of Indian Science Fiction*, Cardiff: University of Wales Press.

Khilnani, S. and Bhattacharjee, R. (2022) *Science Fiction in India: Parallel Worlds and Postcolonial Paradigms*, New Delhi: Bloomsbury.

Koschorke, A. (2018) *Fact and Fiction: Elements of a General Theory of Narrative*, Berlin: De Gruyter.

Kuhad, U. (2022) *Science Fiction and Indian Women Writers: Exploring Radical Potentials*, Abingdon: Routledge.

Lakkad, A.V. (2018) '"Frankenstein's" avatars: posthuman monstrosity in Enthiran/Robot', *Rupkatha Journal on Interdisciplinary Studies in Humanities*, 10(2): 236–250. https://dx.doi.org/10.21659/rupkatha.v10n2.23

Mac Sweeney, N. (2023) *The West: A New History of an Old Idea*, London: W.H. Allen.

Mahalakshmi, M.R. and Rajendran, L. (2023) 'Depiction of science and technology in post-millennial Tamil cinema', *Science, Technology and Society*, 28(4): 639–659.

Ministry of Defence, Government of India (2022a) 'Enhancement of capabilities of AI technology', Press Information Bureau Government of India, press release, 1 August. Available from: https://www.pib.gov.in/PressReleasePage.aspx?PRID=1846937 [Accessed 23 January 2024].

Ministry of Defence, Government of India (2022b) 'Artificial intelligence in defence: presenting AI preparedness of the country in defence', 10 July. Available from: https://www.ddpmod.gov.in/artificial-intelligence-defence [Accessed 23 January 2024].

Ministry of External Affairs of India (2021) 'India's view on the Treaty on the Prohibition of Nuclear Weapons', press release, 22 January. Available from: https://mea.gov.in/press-releases.htm?dtl/33409/Indias+view+on+the+Treaty+on+the+Prohibition+of+Nuclear+Weapons [Accessed 23 January 2024].

Mukherjee, U.P. (2020) *Final Frontiers: Science Fiction and Techno-Science in Non-Aligned India*, Liverpool: Liverpool University Press.

Mukherjee, U.P. (2023) 'From Tafa to Robu: AI in the fiction of Satyajit Ray', in S. Cave and K. Dihal (eds) *Imagining AI. How the World Sees Intelligent Machines*, Oxford: Oxford University Press, pp 237–246.

Nadibaidze, A. (2022) 'Great power identity in Russia's position on autonomous weapons systems', *Contemporary Security Policy*, 43(3): 407–435. https://doi.org/10.1080/13523260.2022.2075665

NITI Aayog (2018) 'National strategy for artificial intelligence', *National Portal of India*, discussion paper, June, Available from: https://niti.gov.in/sites/default/files/2019-01/NationalStrategy-for-AI-Discussion-Paper.pdf [Accessed 23 January 2024].

Onega Jaén, S. and García Landa, J.A. (1999) 'Introduction', in J.S. Onega and J.A. García Landa (eds) *Narratology: An Introduction*, London: Longman, pp 1–41.

Payne, K. (2021) *I, Warbot: The Dawn of Artificially Intelligent Conflict*, London: Hurst & Company.

Qiao-Franco, G. and Bode, I. (2023) 'Weaponised artificial intelligence and Chinese practices of human–machine interaction', *Chinese Journal of International Politics*, 16(1): 106–128. https://doi.org/10.1093/cjip/poac024

Rosendorf, O., Smetana, M. and Vranka, M. (2022) 'Autonomous weapons and ethical judgments: experimental evidence on attitudes toward the military use of "killer robots"', *Peace and Conflict: Journal of Peace Psychology*, 28(2): 177–183. https://doi.org/10.1037/pac0000601

Scharre, P. (2018) *Army of None: Autonomous Weapons and the Future of War*, New York: W.W. Norton.

Scharre, P. (2023) *Four Battlegrounds: Power in the Age of Artificial Intelligence*, New York: W.W. Norton.

Sondhi, J.H. (2022) 'Futurism in Indian cinema: a case study of *Anukul*', in S. Khilnani and R. Bhattacharjee (eds) *Science Fiction in India: Parallel Worlds and Postcolonial Paradigms*, New Delhi: Bloomsbury, pp 208–222.

Suganami, H. (1997) 'Stories of war origins: a narrativist theory of the causes of war', *Review of International Studies*, 23(4): 401–418. DOI: 10.1017/S0260210597004014

Tucker, I. (2023) Signal's Meredith Whittaker: "These are the people who could actually pause AI if they wanted to"', *The Guardian*, 11 June. Available from: https://www.theguardian.com/technology/2023/jun/11/signals-meredith-whittaker-these-are-the-people-who-could-actually-pause-ai-if-they-wanted-to [Accessed 12 June 2023].

Vatsala, P. and Raut, N. (2019) 'Human vision, illusion and disilluson: RA.ONE, a super hero in cyberspace', in N. van Orden (ed) *Navigating Cybercultures*, Leiden: Brill, pp 225–234.

Watts, T. and Bode, I. (2023)' Machine guardians: the Terminator, AI narratives, and US regulatory discourse on lethal autonomous weapons systems', *Cooperation and Conflict*. https://doi.org/10.1177/00108367231198155

Wibben, A.T.R. (2011) *Feminist Security Studies: A Narrative Approach*, Abingdon: Routledge.

II

Technologies and Materialities

ARTWORK

Transformator

Peter Behrbohm, since 2013

Transformator is the first of its kind introducing an all-new executive strategy to the city. Its mission is to refrain from police violence in the near future. After all, it is the human factor that consistently leads to wrong decisions and bloodshed without cause. A central artificial intelligence will therefore decide on the operation of the new fleet of armed machinery that is to be spread all over the city.

Disguised as an electric substation, Transformator is currently extensively tested in the streets of Berlin and Germany as a fully operational autonomous rocket launcher. Passers-by are unsure what is about to happen when a suited gentleman arrives at one of the familiar grey boxes that are to be found all over the city, plugs in the service computer that is built into a suitcase and runs a routine check of the device's active range of motion. The electric box spins around, unfolds its arms in quick movements and suddenly six 44 mm barrels are precisely targeting neighbouring buildings and passing cars. The specialist returns the machine to its hibernation mode, unplugs his computer and leaves the site, heading for the next Transformator on his list.

Artwork details

Peter Behrbohm
Intervention, since 2013, Berlin/Lemgo/Jena
Materials: wood, steel, motors, suitcase, rockets, 27×80×130 cm

Il/legal War: Expanding the Frame of Meaningful Human Control from Military Operations to Democratic Governance

Lucy Suchman

Introduction

Since their formalization in the 19th century, state initiatives in arms control and disarmament have taken as a founding premise the assumption that war is an ineliminable reality of the human condition.[1] According to this logic, one essential means to mitigate violent conflict is through the control of weapons, particularly those designated as unacceptably destructive or inhumane in the suffering that they cause. These include so-called 'weapons of mass destruction' like chemical, biological and nuclear weapons, as well as 'conventional arms' like anti-personnel landmines or cluster munitions that are deemed to kill indiscriminately (and consequently to violate the laws of armed conflict). Even where treaties or agreements encounter resistance from key parties, and despite their widespread violation, the processes of deliberation involved and the resulting frameworks have arguably had significant normative effects.[2]

Without attempting to resolve in principle the question of war's inevitability, this chapter considers another enduring tension: that between militarism and democracy.[3] To develop the argument, I work from a

1 For a historical introduction to arms control and disarmament law, see Casey-Maslen (2022).

2 For the case of nuclear weapons, see Tannenwald (2007); for a wider discussion of the force of international law with respect to war, see Evangelista and Tannenwald (2017).

3 Despite the claims of the US – the world's dominant military power and a state committed to militarism – to be the emblematic democracy, US citizens exercise a demonstrable

central trope in current debates over the regulation of autonomous weapon systems (AWS), the concept of 'meaningful human control'.[4] The frame of meaningful human control has operated as a powerful tool in the context of the debate over AWS and in efforts to challenge the acceleration of weapon systems automation. However, it largely leaves unchallenged the premise of rationality that underwrites modern warfare, remaining within the bounds of the legitimizing logics of command and control. My concern in this chapter is to raise some more fundamental questions regarding the premise that warfighting can be rationally controlled, and to foreground disarmament and demilitarization as alternative paths to security. That means that we begin with the question of war's legitimacy and the realities of impunity in warfighting, then expand the frame of meaningful human control from the operations of weapon systems to wider questions of democratic governance and legal accountability.

I begin with my own attempt, as a participant in discussions at the United Nations Convention on Certain Conventional Weapons (CCW),[5] to set out an argument for the impossibility of developing fully automated weapon systems that could be adherent to international humanitarian law (IHL). Reflecting on the force but also the limits of that argument brings me to the trope of meaningful human control, and the question of how efforts to regulate AWS (pragmatically and perhaps necessarily) frame the bounds of the discussion. I then turn to a key moment in recent history – the US invasion of Iraq in 2003 – to consider how the frame of meaningful human control might be expanded beyond the operation of weapon systems in war to address the question of war's il/legality in particular instances. Drawing on Judith Butler's (2010) theorization of the performative imaginaries of dehumanization that frame war's legitimacy, I conclude with an expanded reading of meaningful human control as forms of democratic governance that might help to shift the balance away from interests vested in the perpetuation of militarism to an expansion of alternatives to war.

lack of democratic control over US foreign policy. I will return to this problem in the concluding discussion.

[4] Wilcox (2023) reminds us, drawing on Black feminist and decolonial scholarship, that the unmarked 'human' is itself a legacy of constitutive histories of dehumanization, racialized violence and colonial control. In the context of warfighting, as Amoore observes, 'the human in the loop is an impossible subject who cannot come before an indeterminate and multiple *we*' (2020: 66, emphasis in original).

[5] The full name is the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects.

1. Briefing the CCW: autonomous (weapon) systems and international humanitarian law

In 2013 the CCW authorized the Chairperson to convene an informal Meeting of Experts to discuss questions related to emerging technologies in the area of lethal AWS.[6] The first informal Meeting of Experts was held in 2014, followed by a further two meetings in 2015 and 2016.[7] As a contribution to the CCW's third informal Meeting of Experts, I presented a briefing paper in a panel focused on the question of autonomy.[8] The paper's argument, which is summarized in this section, was based on my position as an anthropologist engaged for over three decades with the fields of artificial intelligence (AI) and human–machine interaction.[9] My work in tracking developments in AI and robotics involves taking seriously the claims that are made for intelligent machines and comparing them to extensive studies of the competencies – perceptual, and also crucially social and interactional – that are the basis for associated human activities. My focus on situational awareness arises not only from the fact that it is a prerequisite for lawful action within the framework of IHL, but also because this is an area in which I hoped that my perspective could contribute to greater clarity on the concept of autonomy.

1.1 Situational awareness and adherence to the principle of distinction as a necessary condition for lawful autonomy

Designed as a contribution to discussions at the CCW regarding the concept of autonomy, my argument turned on questioning the feasibility of AWS that could comply with IHL.[10] The argument was based less on principle than on practical evidence regarding the interpretive capacities

6 For a timeline, see UNODA (n.d.). Note that the inclusion of the term 'lethal' has been contested by campaigners and some member states, who call for it to be dropped in favour of more inclusive prohibition/regulation of AWS. See Noor, 2023a.

7 Subsequent events included the convening of a Group of Governmental Experts (GGE) that continues to meet. For a critical assessment of these proceedings, see Noor (2023b).

8 See Suchman, 2016.

9 Before taking up my position at Lancaster University, I was a Principal Scientist at Xerox's Palo Alto Research Center, where I spent 20 years as a researcher. In 2002 I received the Benjamin Franklin Medal in Computer and Cognitive Sciences, and in 2010 the ACM SIGCHI Lifetime Research Award. In 1983 I was a founding member of Computer Professionals for Social Responsibility, an organization formed to address the increasing reliance on computing in the control of nuclear weapons systems; I am now a member of the International Committee for Robot Arms Control.

10 International humanitarian law 'is a set of rules which seek, for humanitarian reasons, to limit the effects of armed conflict. It protects persons who are not or are no longer

that legal frameworks like IHL presuppose for their application in a specific situation. These capacities make up what in military terms is called *situational awareness*.[11] Despite other areas of progress in AI and robotics, I offered my assessment that none has been made in the operationalization of situational awareness in an indeterminate environment of action. More specifically for the question of AWS, situational awareness as a prerequisite for the identification and selection of legitimate targets – what has been called the Principle of Distinction[12] – is not translatable into machine-executable code. Yet situational awareness is essential for adherence to IHL or any other form of legally accountable rules of conduct in armed conflict.

1.2 Autonomous weapon systems and the Principle of Distinction

The elements of situational awareness that are most relevant to the question of whether AWS can be adherent to IHL are those that inform *the requirement of distinction* in the use of lethal force; that is, discrimination between legitimate and nonlegitimate targets.[13] In the case of autonomous weapons, adherence to the Principle of Distinction would require that weapon systems have adequate vision or other sensory processing systems, and associated algorithms, for separating combatants from civilians and for reliably differentiating wounded or surrendering combatants from those who pose an imminent threat. Existing technologies such as infrared temperature sensors and associated image processing algorithms may be able to identify something as a human, but they cannot make the discriminations among persons that are required by the Principle of Distinction. Even if machines had adequate sensing mechanisms to detect the difference between civilians and uniform-wearing military, they would fail under situations of contemporary warfare where combatants are most often not in uniform.[14] And more sophisticated technologies such as facial or gait

participating in the hostilities and restricts the means and methods of warfare'. See ICRC Advisory Service, 2004. Campaigners have recently worked to expand the frame to include adherence to international human rights law, which includes injuries outside the context of warfighting.

11 In military doctrine, situational awareness is idealized as 'understanding of the operational environment in all of its dimensions – political, cultural, economic, demographic, as well as military factors' (Dostal, 2001).

12 On the Principle of Distinction, see ICRC IHL Databases (n.d.).

13 I recognize that the requirements of distinction and proportionality are closely linked, but insofar as proportionality judgements presuppose that distinction has been made, I focus on distinction here.

14 Some opponents of a ban on AWS imagine scenarios in which the mere presence of a human body is an adequate criterion for the identification of that person as a legitimate

recognition are still reliant on the existence either of a pre-established database of templates, against which a match can be run, or profiles, which are inherently vulnerable to false positives and other forms of inaccurate categorization.[15]

At this point in the CCW proceedings, there was widespread agreement on a working definition of AWS as weapon systems in which the identification and selection of targets and the initiation of violent force is carried out under machine control; that is, these capacities (identified as 'critical functions') are delegated to the system in ways that preclude deliberative and accountable human intervention. The emphasis in the CCW discussion is specifically on *human* targets; that is, the identification of humans or human-inhabited objects (buildings, vehicles) as lawful targets for engagement. This brackets defensive weapon systems that operate on the basis of unambiguous signals from another (unmanned or uninhabited) device that comprises an imminent threat.

The fundamental problem in meeting the requirements of the Principle of Distinction is that we do not have a definition of a civilian that can be translated into a recognition algorithm. Nor can we get such a definition from IHL. Asaro (2009) reminds us that IHL comprises a diverse body of international laws and agreements (such as the Geneva Conventions), treaties and domestic laws. These are far from algorithmic specifications for decision making and action, as law presupposes in situ forms of judgement regarding its application to a given case. The 1949 Geneva Convention requires the use of 'common sense', while the 1977 'Protocol I' essentially defines a civilian in the negative sense as someone who is not a combatant.[16] Noting

target. But that requirement is counter to the direction in which conflict is moving, as the boundaries that designate geographical zones of combat are increasingly fluid and contested. Moreover, this does not consider the possibility of surrender or other bases for assuming the status *out of combat*.

15 With respect to the development of algorithmic templates for the identification of legitimate targets, Susan Schuppli (2014: 2) observes that 'the recently terminated practice of "signature strikes" in which data-analytics were used to determine emblematic "terrorist" behaviour and match these patterns to potential targets on the ground already points to a future in which intelligence gathering, assessment, and military action, including the calculation of who can legally be killed, will largely be performed by machines based upon an ever expanding database of aggregated information'. The concern here is with an increasing push towards reliance on *a priori* stereotyping rather than systematic intelligence gathering; it is the unreliability of stereotyping that has discredited this practice. A decade later, the developments anticipated by Schuppli have been realized in the Israeli Defense Force's criminal use of algorithmic targeting to accelerate mass killing in Gaza (see Abraham, 2023; Renic and Schwarz, 2023; Gray, 2024). On the exacerbation of the problem of targeted killing by AWS, see also Heyns (2013).

16 Article 50(1) of the Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts, 8 June 1977.

the increasing complexity of the combatant/civilian distinction, Wilke (2014) observes that 'the rise of the figure of the "unlawful combatant" … is accompanied by a corresponding rise of the figure of the illegitimate, noninnocent, suspicious civilian'. Increasingly, in other words, all persons in areas of conflict are liable to be treated as guilty unless proven innocent. While robotics may achieve effective sensory and visual discrimination in certain narrowly constrained circumstances, human-level discrimination with adequate common-sense reasoning for situational awareness would appear to be computationally intractable (Sharkey, 2008). At this point at least, there is no evidence or research result to suggest otherwise.

1.3 The strategic vagueness of law and strategy

The question of autonomy with respect to AWS needs to be considered within a longer history of the intensifying automation of warfare. This is a trajectory justified as a necessary response to the demand for increasingly rapid engagement, along with the vulnerabilities incurred through reliance on complex information and communications networks – a problem that greater automation and system complexity further exacerbates.

We have seen these dynamics before in the implementation of launch on warning in nuclear weapons systems, and some of the questions currently under debate were addressed, and arguably resolved, by the work of computer scientists like David Parnas in the 1980s (Parnas, 1985; see also Smith, 1993). In the context of the US Strategic Defense Initiative, Parnas made the crucial distinction between a computational system's verifiable execution of its specifications on the one hand (what is commonly referred to as the software's 'correctness', or reliability in the narrow sense described in Asaro [2015: 90]) and the system's ability to assess the conditions in which those specifications apply on the other hand (necessary for its reliability in operation). Simulated testing of AWS can assess correctness, but it can never definitively ensure reliability under actual conditions of use. The only way to achieve the latter is through practical methods of iterative development based on repeated trials under conditions that closely match those of intended deployment, or informed by experience of the system in use, neither of which is possible in the case of complex weapon systems with deadly consequences. It was for this reason, among others, that the Strategic Defense Initiative was finally abandoned.

The US Department of Defense *Unmanned Systems Integrated Roadmap 2011–2036* distinguishes automatic from autonomous systems in this passage:

> Dramatic progress in supporting technologies suggests that unprecedented levels of autonomy can be introduced into current and future unmanned systems … Automatic systems are fully

> preprogrammed and act repeatedly and independently of external influence or control … However, the automatic system is not able to define the path according to some given goal or to choose the goal dictating its path. By contrast, autonomous systems are self-directed toward a goal in that they do not require outside control*, but rather are governed by laws and strategies that direct their behavior … The special feature of an autonomous system is its ability to be goal-directed in unpredictable situations*. This ability is a significant improvement in capability compared to the capabilities of automatic systems.[17]

The key phrase here is 'governed by laws and strategies that direct their behavior … in unpredictable situations'. As stated earlier, 'laws and strategies' are not translatable to executable code. In assessing the feasibility of the system posited in this passage, it is crucial to keep in mind that autonomy or 'self-direction' in the case of machines presupposes the unambiguous specification (by human designers) of the conditions under which associated actions should be taken. This requirement for unambiguous specification of condition/action rules marks a crucial difference between autonomy as a human capacity and machine autonomy. As I have argued previously, autonomy as we understand it in the context of human action means self-direction under conditions that are not, and cannot be, fully specified by rule.[18] This in turn accounts for what we might call the *strategic vagueness* of any kind of rule or directive for action; that is, the assumption that the exercise of the rule, or the execution of the directive or plan, will involve in situ judgement regarding the rule's application. Where the requisite competencies are in place, this openness – far from being a problem – is what enables the effectiveness of a general plan or rule as a referent for situated action.

1.4 Limits to information processing as a model of situational awareness

To make this more concrete, we can take the case of human action according to the rules that define military discipline, and, most pertinent to this discussion, IHL and the Principle of Distinction. Because the precise conditions of combat cannot be known in advance, the directives for action in the case of military operations presuppose competencies for their accurate 'execution' that the directive as such does not and cannot fully specify. It is

[17] US Department of Defense, 2011: 43, emphasis added.

[18] See Suchman, 2007. This problem is not resolved by the promises of machine learning to enable derivations from data external to a specified rule, insofar as so-called learning algorithms continue to rely on the availability of datasets rather than open-ended horizons of input (see Asaro, 2015).

for this reason that situational awareness is necessary to effective and, most importantly for our purposes, legally accountable warfare.

Approaches to AI-based robotics share the common requirement that a machine can engage in a sequence of 'sense, think and act' (see Suchman and Weber, 2016: 9–10). However, it is crucial in this context to be wary of the use of evocative terms that anthropomorphize the functionality of computer programs rather than providing technical descriptions of actual capabilities (Sharkey and Suchman, 2013). Does 'sense, think and act' refer to an assembly-line robot that performs an action at a fixed location, in relation to an environment carefully engineered to match its sensing capacities, and where the consequences of failure are nonlethal? Or does it invoke sensing and perception as dynamic and contingent capacities, in open-ended fields of (inter)action, with potentially lethal consequences? In the case of human combatants, the ability to be goal-directed in unpredictable situations presupposes capacities of situational assessment and judgement, in circumstances where the range of those capacities is necessarily open-ended. Moreover, combat situations frequently involve opponents who work hard and ingeniously to identify and defeat any prior assumptions about how they will behave. This is in marked contrast to the situations in which AI techniques, and automation more generally, have been successfully implemented. In any case, the burden of proof here must rest with proponents and require a higher standard than general assertions of progress in AI, which is debatable other than in certain limited technical areas that do not yet begin to address problems of reliable discrimination between legitimate and illegitimate human targets.[19]

1.5 Implications for lawful weapon system autonomy

Citing Article 48 of the First Additional Protocol to the Geneva Conventions, Crootof (2015: 1873) observes that one implication of the Principle of Distinction is that:

> parties are prohibited from using inherently indiscriminate weapons, which are usually defined either as weapons that cannot be directed at lawful targets or as weapons whose effects cannot be controlled. Additionally, any given attack in an armed conflict cannot be

[19] Assertions that 'technology may evolve and meet the requirements [for human target identification] in the future' (cited in United Nations Report of the 2015 Informal Meeting of Experts on Lethal Autonomous Weapons Systems, clause 51a(iii), p. 14), or 'Autonomous technologies *could* lead to more discriminating weapons systems' (clause 51c(iii), p 15, emphasis added) do not comprise evidence-based statements of fact.

indiscriminate: it must be directed at a lawful target and cannot utilize indiscriminate weapons or methods of warfare.

The defining question for AWS is whether the discriminatory capacities that are the precondition for legal killing can be reliably and unambiguously encoded in weapon systems. As noted earlier, this judgement is becoming increasingly difficult for human warfighters, for several reasons. First, the conditions of so-called irregular warfare have removed traditional designations of battle zones and combatants, requiring much more subtle and uncertain readings of the presence of an imminent threat.[20] Second, because military systems involve increasingly complex, distributed, real-time networks of information and communication, the possibilities not only for strategic accuracy but also for noise have increased.[21] And, finally, the intensifying automation of warfare has effected a progressive narrowing of timeframes for situational assessment.

Lawand (2013) proposes that a 'truly autonomous weapon system would be capable of searching for, identifying and applying lethal force to a target, including a human target (enemy combatants), without any human intervention or control'. But in the parenthetical 'enemy combatants' lies the crux of the problem: how is the identification of 'human target' with 'enemy combatant' confirmed? And what uncertainties characterize the category of 'enemy combatant' that confound, rather than clarify, the problem of legitimate target identification in contemporary warfare? Autonomous systems can be made reliable only to the extent that their environments, the conditions of their operation, can be specified, engineered and stabilized; these requirements do not hold in situations of combat.[22] All of the evidence to date indicates that this is at best an unsolved problem for machine autonomy and at worst (and this is my position, for the reasons set out earlier) an unsolvable one.

Concluding my 2016 statement to the CCW, I emphasized that conceptual clarity regarding the capacities that enable situational awareness in the case of

[20] Melissa L. Flagg, Deputy Assistant Secretary of Defense at the US Office of the Undersecretary of Defense for Acquisition, Technology and Logistics research directorate, imagines a situation in which 'a robotic system is in a battle zone, knows the mission, has been thoroughly tested, has the kinetic option, and its communications links have been cut off' and then asks whether that machine should then make the decision to deploy a weapon independently. But it is precisely this clarity that is absent in actual situations of warfighting (see Magnuson, 2016).

[21] On the intransigence of this problem, see, for example, Cronin (2008).

[22] It is widely recognized that 'as the behavior of automated systems becomes more complex, and more dependent on inputs from environmental sensors and external data sources, the less predictable they become' (Asaro, 2015).

human combatants, with a particular focus on the Principle of Distinction, clarifies in turn the requirements for lethal AWS. The defining question for autonomous weapons is whether the discriminatory capacities that are the precondition for legal killing can be reliably and unambiguously encoded. My argument was that they cannot, and that as a consequence, lethal autonomous weapons are in violation of IHL and should be prohibited.

2. The requirement of meaningful human control

While the argument as I presented it in 2016 posited the impossibility of fully automated weapons adhering to the laws of armed conflict, it left open the question of human accountability in the deployment of weapon systems. It is as a way of addressing that question that those campaigning for a ban on autonomous weapons have insisted on the need to preserve 'meaningful human control' over target selection and engagement (Article 36, 2014). In a report issued in February 2016, United Nations (UN) Special Rapporteurs Maina Kiai and Christof Heyns wrote that: 'Autonomous weapons systems that require no meaningful human control should be prohibited' (Kiai and Heyns, 2016). The word 'meaningful' here is meant to anticipate and reject the proposition that any form of oversight over automated target identification constitutes 'human control'.[23] Roboticist and campaigner Noel Sharkey (2014) offers a list of progressively greater levels of human control:

1. human engages with and selects target and initiates any attack;
2. the computer program suggests alternative targets and human chooses which to attack;
3. the program selects target and human must approve before attack;
4. the program selects target and human has restricted time to veto;
5. the program selects target and initiates attack without human involvement.

On Sharkey's analysis, while Level 1, and possibly Level 2, provide for the minimum necessary conditions for meaningful control, the rest do not. It is not only the case that AWS might circumvent meaningful human control; the greater concern is that they could render it impossible. The judgement

[23] For the minimum necessary conditions for meaningful human control, see Article 36, 2013, 2014; Sauer, 2014. Horowitz and Scharre (2015: 4) propose that in its emerging usage, 'meaningful human control has three essential components: Human operators are making informed, conscious decisions about the use of weapons; human operators have sufficient information to ensure the lawfulness of the action they are taking, given what they know about the target, the weapon, and the context for action; and the weapon is designed and tested, and human operators are properly trained, to ensure effective control over the use of the weapon'.

required for effective and legal action according to a rule requires time for the assessment of a current situation, while decreasing timeframes due to increasing automation close down the time available for deliberation. Moreover, the proposed solution of 'human-machine teaming' is only effective to the extent that system designs maintain the system dynamics required (more colloquially, the time humans need) to allow for meaningful human control (Scharre, 2016). This requirement, in turn, poses further limits to weapon system automation.

A final note is that autonomy is best understood not as an individual capacity – whether human or machine – but rather as a capacity enabled by particular configurations of people and technologies. Different configurations make different capacities for action possible. In Suchman and Weber (2016) we developed the argument, implicit in these discussions, that the adjudication of questions of autonomy and responsibility requires as its unit of analysis not the individual person or singular device, but rather weapon systems understood as always *sociotechnical*. As we elaborate, contemporary social theory has effectively challenged the premise that autonomy can be adequately conceptualized as an intrinsic capacity of any entity, whether human or machine. It follows that to understand the agencies of either people or technologies requires an analysis of the dynamics of the relations through which they are conjoined. In thinking about life-critical sociotechnical systems, it is the question of what conditions of possibility a particular configuration affords for human responsibility and accountability that is key.

3. From meaningful human control to democratic governance

While a relational orientation expands the analytic frame of agency and locates it within more extended systems, it also invites further articulation of the historical, political, and economic constituents of sociotechnical systems. With respect to war, the considerations so far in this discussion beg the question of how the military operations in which weapon systems are deployed are themselves mobilized, legitimized and perpetuated. What would it mean to expand the frame of meaningful human control to the operations not only of weapon systems, but also of the actions of military institutions themselves? The relevant humans on this reading are not only those already incorporated into the closed worlds of military operations, but all of us in whose names warfighting is conducted.[24]

[24] Edwards (1996) and Masco (2014) trace the progression of what Edwards calls 'closed world' discourse as it underwrites associated investments in rationality and control through systems engineering, from the Cold War to the so-called War on Terror. For an extension

Shifting the locus of meaningful human control beyond the bounds of the military machine raises questions of democratic governance. What would constitute meaningful human control in the context of this wider frame? What are the requirements for forms of democratic governance capable of holding the perpetrators of war crimes to account, or preventing illegal acts of war from the outset? Once operations are underway, might meaningful human control include refusal to deploy a weapon system at all, given a war's illegality? The question of illegality in warfighting takes us into the realm of geopolitics and the presence (or absence) of multilateral institutions capable of holding governments to account. In the case of the US, investments in hegemonic militarism, economic special interests and associated strategies of geopolitical positioning have historically overridden democratic expression and/or shaped it through dominant discourses. We can take as a case in point the invasion of Iraq in 2003, a protracted military operation with devastating consequences.

On 2 March 2003, the UK's *Observer* newspaper published a front-page story titled 'Revealed: US dirty tricks to win vote on Iraq war' (Bright et al, 2003). The revelation was the result of an unprecedented act of whistleblowing in which Katherine Gun, a translator at the British Government Communications Headquarters (GCHQ), leaked a classified memo sent to members of GCHQ by a senior figure in the US National Security Agency (NSA). Seeking to legitimize an invasion of Iraq in the face of popular opposition and serious questions regarding the legality of such an operation, the NSA memo asked its allied intelligence agency to monitor the private communications of UN delegates for information, personal or otherwise, that could be used to 'give the US an edge' in leveraging support for war through a UN Security Council Resolution.[25] Gun provided the memo to the media with hope that the revelation of the proposed bugging and blackmail tactics might be enough to reverse plans by the Blair government to join the US alliance and interrupt the march to war. Gun confessed to the leak in order to save her GCHQ colleagues from interrogation and was subsequently arrested and charged with breach of the Official Secrets Act. When admonished by a Special Branch officer that 'You work for the British government', Gun famously replied: 'I work for the British people. I do not gather intelligence so the government can lie to the British people'.[26] The case against Gun was dropped by the British

of this line of analysis to the current promise to delineate and dominate the theatre of operations through data, see Suchman (2023).

[25] In 2016 the Chilcot Inquiry revealed that UK government counsel had offered arguments against the legality of the invasion – hence the need for a UN Resolution overriding that assessment. See *The Guardian* Staff, 2016.

[26] Adams, 2019. For selected news reports, see Institute for Public Accuracy (n.d.).

government in 2004 after her lawyer threatened to use disclosure to put the legal basis of the war itself on trial. Gun's planned 'defence of necessity', based on the obligation to act in the face of an imminent threat to life, remained untested.

Regrettably, investment in war by then President George W. Bush, backed by the force of a presentation to the UN of fallacious 'evidence' for weapons of mass destruction, outweighed Gun's revelation and, despite worldwide demonstrations of popular opposition, the invasion proceeded.[27] The case of the invasion of Iraq by the US and its allies, justified by what was suspected at the time to be questionable evidence for an imminent threat and in the face of the most massive worldwide public opposition ever demonstrated, underscores the unequal balance of power between investments in hegemonic militarism and in multilateral institutions of democratic governance. Considered in the context of meaningful human control, this case shifts the focus from control over the operations of a particular weapon system to the prior question of how decisions to initiate or enter into warfighting are adjudicated in the first instance. This suggests that along with multilateral institutions capable of enforcing legal accountability in the conduct of war, meaningful democratic control of the use of force requires accountability for war's initiation and the dismantling of structures of militarism that are invested in war's perpetuation.

Conclusion

In *Frames of War* (2010: xiii) Judith Butler asks: 'do we understand the frame as itself part of the materiality of war and the efficacy of its violence?' As Butler observes, frames are haunted by their constitutive outsides. In the context of Israeli air strikes on Gaza over 22 days in December 2008 and January 2009, Butler observes that: 'The idea of a legal war or, indeed, a just war, relies in the controllability of the instruments of destruction. But because uncontrollability is part of that very destructiveness, there is no war that fails to commit a crime against humanity, a destruction of civilian life' (2010: xviii). The myth of precision, Butler argues, of 'a "clean" war whose destruction has perfect aim' (2010: xviii), is part of the apparatus that holds militarism in place (see also Suchman, 2020). Butler embraces the proposition that a technopolitical imaginary committed to nonviolence requires thinking beyond an instrumentalist framework, asking 'what new possibilities for ethical and political critical thought result from that opening?' (2020: 19). Butler argues that those possibilities require resistance to the charge that

[27] See UN, 2003.

nonviolence is unrealistic, in favour of a critique of what counts as reality and a refusal to be confined within that closed world.

The present chapter, written amid the tragic circumstances of the attack on Israel by Hamas militants in October 2023 and Israel's subsequent, and relentless, assault on Gaza, takes up this larger collection's theme of imaginaries that underwrite investments in weapon systems and war making. My concern is with the question of how prevailing legal and normative debates over the future of war are framed in ways that place more radical approaches to demilitarization off the table. Arms control and disarmament are essential multilateral frameworks within which to mitigate the worst abuses of armed conflict, and the call for meaningful human control in the context of AWS is vital to the project of limiting warfighting's further automation. At the same time, we need to recognize that a technopolitical imaginary of military operations governed by reason and control is part of the problem, not the solution, to relations based on military dominance. AWS are the logical extension of military doctrine that posits further automation of war as the necessary response to warfighting's inevitable acceleration. The result is a self-justifying arms race that benefits those whose returns rely on ever-expanding investments in militarism. The restoration of meaningful control to such a machine requires moving outside of the frame that sustains warfighting's self-justifying logics, enabling greater investments in diplomacy and new forms of accountability to democratic processes, in which those in whose names war is perpetuated would have the governing voice.

References

Abraham, Y. (2023) '"A mass assassination factory": inside Israel's calculated bombing of Gaza', *+972 Magazine*, 30 November. Available from: https://www.972mag.com/mass-assassination-factory-israel-calculated-bombing-gaza/ [Accessed 4 December 2024].

Adams, T. (2019) 'Iraq war whistleblower Katharine Gun: "truth always matters"', *The Observer*, 22 September, Available from: https://www.theguardian.com/film/2019/sep/22/katharine-gun-whistleblower-iraq-official-secrets-film-keira-knightley [Accessed 20 July 2023].

Amoore, L. (2020) *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*, Durham, NC: Duke University Press.

Article 36 (2013) 'Killer robots: UK government policy on fully autonomous weapons', April, Available from: http://www.article36.org/wp-content/uploads/2013/04/Policy_Paper1.pdf [Accessed 20 July 2023].

Article 36 (2014) 'Key areas for debate on autonomous weapon systems: memorandum for delegates at the Convention on Certain Conventional Weapons', *Meeting of Experts on Lethal Autonomous Weapons Systems*, 13–16 May. Available from www.article36.org/wp-content/uploads/2014/05/A36-CCW-May-2014.pdf [Accessed 21 December 2023].

Asaro, P. (2009) 'How just could a robot war be?', in P. Brey, A. Briggle and K. Waelbers (eds) *Current Issues in Computing And Philosophy*, Amsterdam: IOS Press, pp 50–64.

Asaro, P. (2015) 'Roberto Cordeschi on cybernetics and autonomous weapons: reflections and responses', *Paradigmi: Rivista di critica filosofica*, XXXIII(3): 83–107.

Bright, M., Vulliamy, E. and Beaumont, P. (2003) 'Revealed: US dirty tricks to win vote on Iraq war', *The Guardian*, 2 March. Available from: https://www.theguardian.com/world/2003/mar/02/usa.iraq [Accessed 20 July 2023].

Butler, J. (2010) *Frames of War: When Is Life Grievable?*, London: Verso.

Butler, J. (2020) *The Force of Non-Violence*, London: Verso.

Casey-Maslen, S. (2022) *Arms Control and Disarmament Law*, Oxford: Oxford University Press.

Cronin, P. (2008) *The Impenetrable Fog of War: Reflections on Modern Warfare and Strategic Surprise*, Westport, CT: Praeger Security International.

Crootof, A. (2015) 'The killer robots are here: legal and policy implications', *Cardozo Law Review*, 36: 1837–1915.

Dostal, B.C. (2001) 'Enhancing situational understanding through the employment of unmanned aerial vehicles', *Center for Army Lessons Learned*. Available from: http://www.globalsecurity.org/military/library/report/call/call_01-18_ch6.htm [Accessed 23 July 2023].

Edwards, P. (1996) *The Closed World: Computers and the Politics of Discourse in Cold War America*, Cambridge, MA: MIT Press.

Evangelista, M. and Tannenwald, N. (2017) *Do the Geneva Conventions Matter?*, Oxford: Oxford University Press.

Gray, C.H. (2024) 'Military AI, sacred violence and war in the Middle East', *Teknokultura*, 21(2): 167–171.

The Guardian Staff (2016) 'Chilcot report: key points from the Iraq inquiry', *The Guardian*, 6 July. Available from: https://www.theguardian.com/uk-news/2016/jul/06/iraq-inquiry-key-points-from-the-chilcot-report [Accessed 23 July 2023].

Heyns, C. (2013) 'UN Doc. A/HRC/23/47, Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions', *United Nations*, 9 April. Available from: https://documents-dds-ny.un.org/doc/UNDOC/GEN/G13/127/76/PDF/G1312776.pdf?OpenElement [Accessed 23 July 2023].

Horowitz, M. and Scharre, P. (2015) 'Meaningful human control in weapon systems: a primer', *Center for a New American Security*, March. Available from: https://www.jstor.org/stable/pdf/resrep06179.pdf [Accessed 23 July 2023].

ICRC Advisory Service (2004) 'What is international humanitarian law?', *ICRC Advisory Service*, Available from: https://www.icrc.org/eng/assets/files/other/what_is_ihl.pdf [Accessed 20 July 2023].

ICRC IHL Databases (n.d.) 'International humanitarian law databases, I. The principle of distinction', *ICRC IHL Databases*, Available from: https://ihl-databases.icrc.org/en/customary-ihl/v1 [Accessed 21 December 2023].

Institute for Public Accuracy (n.d.) 'The Katherine Gun case', *Institute for Public Accuracy*. Available from: https://web.archive.org/web/20040401190525/http://www.accuracy.org/gun/ [Accessed 21 December 2023].

Kiai, M. and Heyns, C. (2016) 'Joint report of the Special Rapporteur on the rights to freedom of peaceful assembly and of association and the Special Rapporteur on extrajudicial, summary or arbitrary executions on the proper management of assemblies', 4 February, A/HRC/31/66. Available from: http://freeassembly.net/reports/managing-assemblies/ [Accessed 23 July 2023].

Lawand, K. (2013) 'Fully autonomous weapon systems', *International Committee of the Red Cross*, 25 November. Available from: http://www.icrc.org/eng/resources/documents/statement/2013/09-03-autonomous-weapons.htm [Accessed 20 July 2023].

Magnuson, S. (2016) 'Autonomous, lethal robot concepts must be "on the table", DoD official says', *National Defense Magazine*, 3 March, Available from: https://www.nationaldefensemagazine.org/articles/2016/3/2/autonomous-lethal-robot-concepts-must-be-on-the-table-dod-official-says [Accessed 20 July 2023].

Masco, J. (2014) *The Theater of Operations: National security from the Cold War to the War on Terror*, Durham and London: Duke University Press.

Noor, O. (2023a) 'States make progress on policy at UN discussions', *Stop Killer Robots*, 14 March, Available from: https://www.stopkillerrobots.org/news/states-make-progress-on-policy-at-un-discussions-as-momentum-builds-towards-treaty-on-aws/ [Accessed 20 July 2023].

Noor, O. (2023b) 'Convention on Conventional Weapons runs out of road as states adopt meaningless report', *Stop Killer Robots*, 22 May. Available from: https://www.stopkillerrobots.org/news/states-adopt-meaningless-report-after-civil-society-excluded-from-un-discussions-on-autonomous-weapons-systems/ [Accessed 20 July 2023].

Parnas, D.L. (1985) 'Software aspects of strategic defense systems', *Comm ACM*, 28 (12): 1326–1235.

Renic, N. and Schwarz, E. (2023) 'Inhuman-in-the-loop: AI-targeting and the erosion of moral restraint', *Opinio Juris*, 19 December. Available from: https://opiniojuris.org/2023/12/19/inhuman-in-the-loop-ai-targeting-and-the-erosion-of-moral-restraint/ [Accessed 16 August 2024].

Sauer, F. (2014) 'ICRAC statement on technical issues to the 2014 UN CCW Expert Meeting', *ICRAC*, 14 May. Available from: https://www.icrac.net/icrac-closing-statement-to-the-2014-un-ccw-expert-meeting/ [Accessed 23 July 2023].

Scharre, P. (2016) 'Autonomous weapons and operational risk', *Center for a New American Security*, February. Available from: https://www.stopkillerrobots.org/wp-content/uploads/2021/09/CNAS_Autonomous-weapons-operational-risk.pdf [Accessed 20 July 2023].

Schuppli, S. (2014) 'Deadly algorithms: can legal codes hold software accountable for code that kills?', *Radical Philosophy* 187. Available from: https://www.radicalphilosophy.com/commentary/deadly-algorithms [Accessed 20 July 2023].

Sharkey, N. (2008) 'Grounds for discrimination: autonomous robot weapons', *RUSI Defence Systems*, 11(2): 86–89.

Sharkey, N. (2014) 'Towards a principle for the human supervisory control of robot weapons', *Politica & Societa*, 3(2): 305–324.

Sharkey, N. and Suchman, L. (2013) 'Wishful mnemonics and autonomous killing machines', *AISBQ Quarterly*, 136: 14–22.

Smith, B.C. (1993) 'Limits of correctness in computers', in T.R. Colburn, J.H. Fetzer and T.L. Rankin (eds) *Program Verification. Studies in Cognitive Systems*, vol 14, Dordrecht: Springer, pp 275–293.

Suchman, L. (2007) *Human-Machine Reconfigurations*, New York: Cambridge University Press.

Suchman, L. (2016) 'Situational awareness and adherence to the principle of distinction as a necessary condition for lawful autonomy', in R. Geis (ed) *Lethal Autonomous Weapon Systems*, Berlin: Federal Foreign Office, pp 273–283.

Suchman, L. (2020) 'Algorithmic warfare and the reinvention of accuracy', *Critical Studies on Security*, 8(2): 175–187. https://doi.org/10.1080/21624887.2020.1760587

Suchman, L. (2023) 'Imaginaries of omniscience: automating intelligence in the US Department of Defense', *Social Studies of Science*, 53(5): 761–786. https://doi.org/10.1177/03063127221104938

Suchman, L. and Weber, J. (2016) 'Human-machine autonomies', in N. Bhuta, S. Beck, R. Geis, H.-Y. Liu, and C. Kreis (eds) *Autonomous Weapons Systems*, Cambridge: Cambridge University Press, pp 75–102.

Tannenwald, N. (2007). *The Nuclear Taboo: The United States and the Non-use of Nuclear Weapons since 1945*, Cambridge: Cambridge University Press.

UN (United Nations) (2003) 'Briefing Security Council, US Secretary of State Powell presents evidence of Iraq's failure to disarm', 5 February, Available from: https://press.un.org/en/2003/sc7658.doc.htm [Accessed 20 July 2023].

UN (2015) 'Report of the 2015 Informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS)', *United Nations*. Available from: https://documents-dds-ny.un.org/doc/UNDOC/GEN/G15/111/60/PDF/G1511160.pdf?OpenElement [Accessed 21 December 2023].

UNODA (United Nations Office for Disarmament Affairs) (n.d.) 'Timeline of LAWS in the CCW'. Available from: https://disarmament.unoda.org/timeline-of-laws-in-the-ccw/ [Accessed 23 December 2023].

US Department of Defense (2011) 'Unmanned Systems Integrated Roadmap FY2011-2036'. Available from: https://irp.fas.org/program/collect/usroadmap2011.pdf [Accessed 2 January 2025].

US Department of Defense (2012) 'Directive 3000.09: autonomy in weapon systems', 21 November 2012'; updated 25 January 2023. Available from: https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf [Accessed 20 July 2023].

Wilke, C. (2014) 'Civilians, combatants, and histories of international law', *Critical Legal Thinking*, 28 July. Available from: http://criticallegalthinking.com/2014/07/28/civilians-combatants-histories-international-law/ [Accessed 20 July 2023].

Wilcox, L. (2023) 'No humans in the loop: killer robots, race, and AI', in J. Brown et al (eds) *Feminist AI: Critical Perspectives on Algorithms, Data, and Intelligent Machines,* Oxford: Oxford University Press, pp 83–100.

From Network-Centric Warfare to Autonomous Warfighting Networks: Recontextualizing Autonomous Weapon Systems Imaginaries

Christoph Ernst

Introduction

Like many discussions about new technologies, the debate about autonomous weapon systems (AWS) is shaped by collective imagination. Imaginative processes contribute significantly to how a society evaluates the effects of a technology and what specific possibilities are associated with it. Whether in more technical discourses about the distinctive capabilities of a certain AWS, academic debates about the autonomy of these weapons or speculative considerations about the consequences of AWS for future warfare, collective processes of imagination – so-called imaginaries – are always at work when different social actors form an 'image' of the future possibilities of this technology.

Imaginaries are a subject of research in science and technology studies (STS), sociology and philosophy of technology, and media studies. Concepts such as 'sociotechnical imaginaries' (Jasanoff, 2015; McNeil, 2017), 'techno futures' ('Technikzukünfte') (Grunwald, 2012) or 'media prophecies' (Natale and Balbi, 2014: 205–207) are used for their analysis. These terminologies target the fact that collective imaginaries are accompanied by political assessments of the social consequences of a technology. In order to recognize underlying structures and recurring patterns, research on imaginaries is in large parts historical. The historical analysis and re-appraisal of past futures is a field closely related to research on imaginaries. Given the central role of warfare in the relation between society and technology, imaginaries of coming weapons systems and future forms of warfare have always played an integral role in this research.[1]

Following this line of research, the argument of this chapter focuses on the idea of network-centricity as a crucial element within the context of current imaginaries associated with AWS. It is argued that ideas on so-called network-centric warfare (NCW) developed during the 1990s and early 2000s are important historical sources which provide scripts and metaphors for contemporary AWS imaginaries. Analysing the significance of the transformation of those imaginaries from the 1990s and 2000s towards a more current concept of network-centric warfare, it becomes apparent that while there is much discussion about the importance of artificial intelligence (AI) for warfare, the relevancy of network-centricity for AWS imaginaries and the associated visions of future warfare is often overlooked. Recent AWS imaginaries contain the infrastructural vision of what can be called autonomous warfighting networks. It will be argued that imaginaries of AI, and by implication AWS, cannot be understood without highlighting these ideas and visions surrounding the contemporary forms of network-centric warfare.

In order to do so, the first section provides a systematic exploration of research on imaginaries relevant for the debate on AWS. The term 'imaginaries' will be clarified and contextualized within discussions on AWS imaginaries. The second section takes a look back at some of the core ideas of network-centric warfare, especially in the early 2000s. These ideas revolved around the idea of the network as a superior form of information organization. Yet, with the advent of the internet of things, the understanding of 'network' changed on an infrastructural level. This shift is reflected in the third section. As expressed in the notion of 'combat clouds', the network – defined as an entity consisting of relations between different material entities – is at the heart of current imaginaries of AWS. While the discussion on AWS leans heavily towards questions of human control of autonomous systems, the analysis of the respective imaginaries shows the emergence of a different scenario: prolonging ideas of network-centricity AWS are imagined in current debates as autonomous warfighting networks.

1. Systematic aspects of autonomous weapon systems imaginaries

Current research on imaginaries in science and technology-studies (STS) is heavily informed by Sheila Jasanoff's (2015) concept of 'sociotechnical

[1] See, among others, Corn and Horrigan, 1984: 108–134; Freedman, 2018: 222–287.

imaginaries'.[2] In the context of societal imaginations of technology, imaginaries are associated with the assumption that technology is a primary formative force of social change. They create images, scenarios or narratives of future technological possibilities within the context of a techno-political goal. Imaginaries are thus to be understood as elaborations of expected possibilities of a technology which are institutionally supported within the context of political and economic purposes. Through imaginaries, technological possibilities are illustrated in terms of their consequences for future ways of life (Jasanoff, 2015: 4). Being socially produced, imaginaries are therefore specific, historically variable and changeable figurations.[3]

Most of the time, a strong focus of an imaginary is on scenarios that can be expected in the near future – that is, the imagination of possibilities that are linked to a real technology (for example, in the context of media technologies, see Ernst and Schröter [2021]). Usually, imaginaries refer to technologies that already have a certain degree of development and a high promise of innovation in the sense of a fundamental transformation of societal conditions.[4] This explains why imaginaries are found especially in contexts where the possibilities associated with a technological innovation become political. Due to some technological or political development, they are considered to be *realistic* possibilities. Imaginaries then serve to manage expectations, promises and forecasts concerning such 'new' possibilities. They create an image of the (near) future by connecting a technology with generally comprehensible scenarios for its application.[5]

---

[2] Jasanoff's approach stands in a longer tradition of philosophical considerations on the societal functions of imagination. Most eminent are the works by Cornelius Castoriadis (1987), which elaborate the idea that society itself is a product of imagination. Furthermore, there is a broad sociophilosophical debate on the social function of imaginaries that goes far beyond specific 'sociotechnical' contexts (Taylor, 2004).

[3] Imaginaries are not to be equated with the human capacity for imagination per se, or with the cultural imaginary as such. One can think, for example, of the imagination of the atomic bomb. For Günther Anders (1994 [1956]), the discussion of the atomic bomb is linked to the problem that the consequences of these weapons could, in the final analysis, not be imaginable at all – that is, they touch the limits of imaginability as such. Yet, this does not mean that there are no imaginaries surrounding nuclear weapons. But from Anders' point of view, these imaginaries would always be distorting and thoroughly ideological. Thus, Anders highlights the difference between imagination, the (cultural) imaginary, and political imaginaries in particular (for further context, see Ernst and Schröter, 2021: 29–35).

[4] The envisioned technological transformation in almost all cases appears with regard to fields of society in which technological change has a long tradition such as mobility, energy, communication, health, the economy and warfare. An example would be quantum computing as a fundamentally new principle within the long tradition of computing – see Schröter et al (2022).

[5] Accordingly, imaginaries stand in relation to other patterns of imagination, in particular utopias and dystopias, as they are part of the cultural imaginary. For utopian and dystopian

Imaginaries are an integral part of political communication because they consist of programmatic or prototypical statements that address heterogeneous audiences. Thus, they are able to perform 'boundary work' (Gieryn, 1983). Their function is to coalesce a multitude of different interests and knowledge horizons and familiarize various actors with a common vision of particular futures associated with a technology (Flichy, 2007: 7–13). Hence, the primary habitat of imaginaries are the wider and more heterogeneous public spheres of large institutions (that is, the public services or armed forces of a nation state as a whole) as well as the multitude of different public sphere(s) of a given society as a whole.[6]

In light of this systematic scope of the term and the complexities associated with it, the question arises of how an imaginary can be identified and delimited. Regarding this issue, one has to acknowledge that the pragmatic functions of imaginaries are more important than the homogeneity of their discursive realization and the differentiability of distinct semantic features. In other words, imaginaries can serve pragmatic purposes *without* necessarily having to be clearly explicable. They work on an informal or tacit level. This is associated with their highly transversal character. Imaginaries are neither bound to certain modes of articulation (fiction, factual and so on) nor to specific genres (science fiction, fantasy and so on). Given their reference to the future, imaginaries contain a high proportion of fictional, speculative, hypothetical or counterfactual statements. They can be fictional from the outset as well, for example, in the context of 'diegetic prototypes' (Kirby, 2010). Furthermore, imaginaries can appear in all known media and formats. In addition to written, pictorial or diagrammatic mediations, practical-performative articulations such as technology demonstrations (Grunwald, 2012; Rosental, 2021; Eckel et al, 2023) or public scientific experiments (Smith, 2009) have to be considered as well.

This perspective can be easily applied on AWS. In a brief overview of different forms of contemporary AWS imaginaries, Jutta Weber (2021) distinguishes three heterogeneous contexts in which imaginaries of these weapons systems currently appear:

1. Fictional or documentary formats that pursue a politically enlightening purpose, such as the well-known 'Slaughterbots' video by the Future

views on artificial intelligence, see, for example, Cave and Dihal (2019) and, in a broader perspective on AI-narratives in general, Cave, Dihal and Dillon (2020).

[6] In the contexts of highly specialised professional communities, which are concerned with problem solving on a very practical level, imaginaries are only relevant because they influence prevailing 'Leitbilder' ('guiding images') of technological development and problem solving (Hellige, 1992).

of Life Institute (2017).[7] From the perspective of a spokesperson of a hypothetical nongovernmental organization (NGO), this fictional video warns of the dangers of AWS by illustrating the attack of a semi-autonomous drone swarm on civilian targets (Weber, 2021: 170–171).
2. Publications by the military or the defence industry that illustrate the capabilities of autonomous systems. Weber discusses a video by the US Navy entitled 'Perdix Swarm Demonstration' (US Navy, 2017), which has also become quite prominent. It shows the test of the autonomous operations of a swarm of Perdix drones over China Lake in California in 2016 (Weber, 2021: 171–172).
3. Films and series – Weber mentions exclusively Hollywood productions, which have shaped the imagination of AI at least since classics such as *2001: A Space Odyssey (*1968) or *Ex Machina* (2015). In the context of AWS, of course, the *Terminator* franchise (1984–2019) is most often referenced (Weber, 2021: 172–175).

All these aspects mentioned by Weber are part of a long tradition of collective imagination of possibilities associated with weapon systems of all kinds in the 19th and 20th centuries (for example, Gannon, 2003). The mere fact that in the first half of the 21st century, with AWS, a new class of weapon systems became the focus of collectively negotiated and publicly articulated imaginaries of their future possibilities is certainly not unusual. It is nevertheless important to understand how AWS imaginaries are, following Jasanoff, 'institutionally stabilized' (Jasanoff, 2015: 4) in current discourse. This brings the political dimension of imaginaries to the fore.

As Weber (2021: 167–170, 175) highlights, the futures articulated in AWS imaginaries are politically contested. AWS imaginaries are (often) characterized by an interweaving of fact and fiction and pursue different political goals. For example, they are linked to the history of the imagination around AI and follow prominent metaphors, for example, that of the swarm (Arquilla and Ronfeldt, 2000). Depending on the political goal, different interpretations of these metaphors emphasize certain capabilities and undercut others.[8] From this emerges their relationship to other forms of imagination of the future within the general cultural imaginary. While AWS imaginaries do consist of explicit references to commonly understood ideas around the consequences of autonomous (weapons) systems which are easily readable and understandable, what they communicate is a different story. An actual imaginary consists primarily of subtextual or connotative

7 In the meantime a sequel has been published: see Future of Life Institute (2021).
8 In fictional contexts, their place is therefore in what Timothy Lenoir and Luke Caldwell (2018) have very convincingly analysed as the 'Military-Entertainment Complex'.

evocations, for example, of common collective symbols, topoi and metaphors that are anchored in cultural memory. When AWS imaginaries formulate promises, evoke expectations or stir up fears with regard to technological upheavals that follow from the automatization of weapons systems, they gain their persuasive power from implicit cross-references to more deeply rooted utopias and dystopias. Hence, the primary mode of operation of an imaginary works on an informal or tacit level of communication.

A helpful approach that describes this transition between the explicit sides of imaginaries and the deeper structures of the cultural imaginary can be found in media studies. For the second half of the 20th century, three so-called 'cores of fascination' ('Faszinationskerne') for new media technologies have been identified (Glaubitz et al, 2011; Ernst and Schröter, 2021: 43–49). The three cores identified in this research are 'simulation', 'networks/networking' and 'artificial intelligence'.[9] Following the authors, the cores of fascination form a bridge between historically deeply rooted ideas and much more recent imaginaries. For example, in the case of AI as a core of fascination, a line can be traced back to imaginaries around the loss of human autonomy and the image of enslavement by machines (Cave and Dihal, 2020; Dihal, 2020). Incarnations of this idea in well-known blockbuster productions like *The Matrix* (US, 1999) or the television series *Westworld* (US, 2016–2022) merge motifs around AI with ideas on simulation, illustrating the claimed interconnectedness of the three cores of fascination. The idea of such cores of fascination within AWS imaginaries can be used to take a closer look at ideas and promises regarding networks in recent AWS imaginaries.

2. The long shadow of network-centric warfare

The conception of AWS plays a central role in current imaginaries on future warfare in general. Since the end of the 2000s, drones have been emblematic in AWS imaginaries. Recently, however, the public representation of AWS has increasingly focused on the interaction of manned and unmanned systems. A very prominent public example is the so-called FCAS project. FCAS stands for 'Future Combat Air System', a 6th-generation combat aircraft being developed by Germany and France. On 28 April 2023, *Tagesschau*, one of the major German-language television news programmes, reported that Spain had decided to join Germany and France:

[9] Glaubitz et al (2011: 143–162) frame AI as 'medial actors' (*mediale Aktanten*), highlighting, among other aspects, the relevancy of robotics within AI imaginaries. Various 'media prophecies' (Natale and Balbi, 2014) that emerged around digital media in the 20th century gravitate around such cores of fascination.

> An estimated 100 billion euros are flowing into the development of FCAS. This makes it the most expensive European armament project to date. The new fighter jet alone is not enough: FCAS is to be a weapons system that also includes drones, artificial intelligence and satellite technology. It is about networked warfare: a fighter jet is in direct connection with a swarm of drones and perhaps also with warships at sea. (Neuroth, 2023, author's translation)

FCAS is the central part of a 'Next Generation Weapon System' (NGWS), the goal of which is claimed to be 'networked warfare'. According to this, networked warfare is to be understood as the 'direct connection' between a manned combat aircraft and a 'swarm' of unmanned drones and other actors such as naval assets (Neuroth, 2023, author's translation). Further details are provided by other publications. For example, the magazine *Behörden Spiegel*, whose target audience are public service employees in Germany, states in a special issue on this 6th-generation combat aircraft:

> In the development of the sixth-generation air combat systems, a combination of manned and unmanned flight systems into a single unit will be realised for the first time … The combat crew then no longer 'only' controls and operates their aircraft, they lead a swarm of unmanned remote carriers with different capabilities in the field of reconnaissance, kinetic and non-kinetic effects and possibly also in command and control support. How this 'manned-unmanned-teaming' interaction is supposed to work has been part of in-depth research for decades. 'The more complex-the more manned. The more dangerous-the more unmanned' – this could be a line of action for the division of tasks. The challenges of making this system network fully effective in a resilient manner, even at the highest combat intensity, are immense. The sixth-generation combat aircraft (as the F-35 system already demonstrates) require a data network as a 'system of systems' that transports all relevant command and reconnaissance information in the dimensions of land, air, sea, cyber and space with low latency and high bandwidth in the subsonic and supersonic range through combat cloud solutions. (Wolski, 2022: 4, author's translation)

What makes the quotes noteworthy is its subtextual historical perspective. The 'networked warfare' that is mentioned aims at the interaction of manned and unmanned systems. Such an understanding of networked warfare as a 'manned-unmanned teaming' has in recent years overshadowed the older understanding of 'networking' and the 'network' as an effective way to distribute information. However, the implementation of a 'combat cloud' – that is, a network in which there is a constant exchange of information with central servers – is also mentioned. The combat cloud is seen as a necessary condition, a prerequisite for the concept of manned-unmanned teaming and the ideal typical division of labour between autonomous systems and humans which is derived from it. Autonomous systems are imagined to do relatively easy but highly risky tasks, while humans remain in the background, monitoring and assessing the situation in the battlespace.

In the 1990s and early 2000s, there was also talk of 'networked warfare', albeit under very different circumstances. The catchword network-centric warfare (NCW) was an echo of the prominence of the network-metaphor at the time. In 1996 Manuel Castells published his well-received book *The Information Age: Economy, Society, and Culture, Volume 1: The Rise of the Network Society* (Castells, 1996). Accordingly, NCW was supposed to be a concept for warfare within the framework of the new societal conditions described by Castells and others (Guha 2011: 86–132). NCW was a transfer of ideas on the use of networked computers from media culture and economics to warfare and was developed and forcefully promoted by the US think-tank culture, especially the RAND Corporation. A few of the most important books from this period make the basic idea clear in their titles: *Understanding Information Age Warfare* (Alberts et al, 2001) and *Network Centric Warfare. Developing and Leveraging Information Superiority* (Alberts, Garstka and Stein, 2001).[10]

In principle, the discussion on reimagining future warfare by thinking about 'robotic wars' had gained momentum in the 1980s and after the 1991 Gulf War (de Landa, 1991: 127–178; Gray, 1997; Singer, 2010). Yet, the focus of NCW was different. At its centre was the idea of rethinking information as an independent variable in warfare in conceptual and infrastructural terms. NCW was understood as a concept for achieving 'information superiority' (Alberts, Gartska and Stein, 2000: 54–58; Alberts et al, 2001: 1–2). The declared aim was not simply to 'network' different military systems, but to let them *act* as a network in the sense of an entity superior to other ways of organizing relations between different platforms of weapon systems (Alberts, Gartska and Stein, 2000: 115–131; Alberts et al, 2001: 49–50). Networked information and communication technologies (ICTs) were proposed as a 'system of systems' (Guha 2011: 93–97). The superiority of this decentralized metasystem would result from the advantages in information distribution and allocation that networks have over other organizational forms.

In 2001 the US Department of Defense established the so-called 'office for force transformation'. Vice Admiral Arthur Cebrowski, its first director, was the most prominent propagator of NCW in the US military. In late

[10] Other important sources for the understanding of NCW during this era are Arquilla and Ronfeldt (2000) and Cebrowski (2005).

2003, his office released a comprehensive overview on the advantages of NCW. The overview illustrates the supposed benefits of this new 'systems-of-systems' well. It states as follows:

> What is Network-Centric Warfare?
>
> The term 'network-centric warfare' broadly describes the combination of emerging tactics, techniques, and procedures that a fully or even partially networked force can employ to create a decisive warfighting advantage.
>
> - NCW is an information superiority-enabled concept of operations that describes the way U.S. forces organize and fight in the information age.
> - NCW generates increased combat power by networking sensors, decision makers, and shooters to achieve shared awareness, increased speed of command, high tempo of operations, greater lethality, increased survivability, and a degree of self-synchronization.
> - NCW translates information superiority into combat power by effectively linking friendly forces within the battlespace, providing a much improved shared awareness of the situation, and enabling more rapid, effective decision making. (Department of Defense, 2003: 3)

According to this, information superiority is, among other factors, archived by connecting sensors, commanders and warfighters in one meta-network. This is supposed to create synergies leading to 'shared awareness', meaning a better mental concept of what is going on in a 'battlespace'. Furthermore, claims regarding superior speed in decision making and warfighting are associated with the idea of 'self-synchronization'. Self-synchronization describes, for example, how through the use of networked information about the consumption of fuel, ammunition and so on, supplies are always already in place without the necessity of explicitly requesting support by the fighting force (Alberts, Gartska and Stein, 2000: 175–180). Their need for logistical support is already anticipated by the system, enabling the fighting force to stay in action seamlessly.

As already indicated, such ideas to regard network-centricity as the future of warfare did not materialize out of thin air. The discourse on the 'information age' and 'network society', from which NCW was derived, was heavily informed by and oriented towards the then-evolving media technology of the internet and associated terms like William Gibson's metaphor of the 'cyberspace' (Alberts et al, 2001: 46). However, around a quarter of a century later, the meaning associated with 'networked warfare' has obviously changed considerably compared to the early 2000s. Most important is the idea of the 'Internet of Things' (IoT), which appeared in 2004 (Gershenfeld et al, 2004; Sprenger and Engemann, 2015). From

2010 onwards, IoT technologies became elementary prerequisites for new applications of autonomous systems, such as autonomous cars (Sprenger, 2021). With the expansion of the internet, social networks and the IoT, AI in the form of deep neural networks experienced a dramatic upswing, evolving specifically in debates around the problem of 'automation' and 'human control' (Russell, 2019: 103–131). As a consequence, the vision of the early 2000s of networking as a form of interfacing between different weapons systems for the purpose of information exchange has since not disappeared, but has faded into the background. In the now-realized age of 'ubiquitous computing' (Weiser, 1991), the availability and presence of complex ICT networks is seen as a normalized fact. Given that 'combat clouds' still form a central element of AWS imaginaries, it is important to take a closer look at the motif of 'networked warfare' within AWS imaginaries.

3. Combat clouds: autonomous warfighting networks

Autonomy in weapon systems is widely defined by the AI-enabled ability of weapons systems to select, attack and engage targets without human intervention (Bhuta et al, 2016). Because this capability is one that has been acquired through developments in the field of AI and implemented in robotic systems, AWS imaginaries have obvious intersections with the tradition of imaginaries about the social consequences of AI. Particularly in the debates on 'meaningful human control', the motif of human 'obsolescence' (Cave and Dihal, 2019) plays a central role in the face of AI-enabled systems. However, looking at the preceding quotes from the context of the FCAS project, it would be a mistake to discuss AWS imaginaries one-sidedly via AI as the sole core of fascination for imaginaries.

Even though media is full of cross-references between AWS imaginaries and AI, ideas surrounding the network should not be ignored as a core of fascination within AWS imaginaries. Already in the passages quoted previously, the 'combat cloud' – as a network – is explicitly named as a necessary condition for AI and thus for 'autonomy' in weapons systems. This central role of the network as a source of various concepts for contemporary solutions of information exchange often remains in the background in AWS imaginaries. Their focus is either on the integration of human and machine or an antagonism between human and machine – a problem obviously implied in scenarios of manned-unmanned teaming and the associated distribution of tasks and labour between humans and machines. Given the technological landscape, the fading of the network into the background of current AWS imaginaries is, of course, no surprise. What was still imagined as the (near) future around 2000 in the context of the NCW is now our infrastructural reality. Accordingly, the focus has shifted to the material expansion of the network through the integration of AI-based unmanned systems that supposedly act autonomously. Nevertheless, what discussions on AWS imaginaries somewhat have lost sight of is that the concepts surrounding the envisioned combat clouds and comparable networking projects are still praising information superiority and superior situational awareness with nearly similar euphoria as they were in around 2000 (Ernst, 2018).

In the case of the FCAS project, Airbus, Thales and Indra are involved in the development of the envisioned combat cloud solutions. Using the slogan 'Future Combat Mission System', a consortium of four large German defence companies (FCMS GbR, 2023) also plays an important role.[11] Corresponding project names are 'Multi-Domain Combat Cloud' (Airbus) or '4π Meta Sensor Effector System (MSES)' (FCMS GbR, 2023). In the self-portraits of these projects in advertising brochures or videos, the key component of future combat clouds is the networking of sensors, that is, the main source of current data that is shared and evaluated in the network. Continuing an idea from the early 2000s, the aim is to optimize the sensor-to-shooter-circle, that is, the relationship between the acquisition of information through media and the use of effectors (actual weapons) (Ernst, 2018: 420–424). Combat clouds, we learn, serve to transform 'data into actionable information thanks to the latest analytical and learning technologies' (Airbus, 2020) – and hence AI. As Airbus emphasizes, AI-based methods such as 'activity-based intelligence' (Biltgen and Ryan, 2016) play a crucial role in this. As the name of the Airbus project already suggests, it does not matter whether it is about air warfare, as with FCAS, or another domain (land, sea, space or cyber). The solution fits every aspect of warfare: 'Multi-Domain Combat Cloud enables collaborative combat with manned and unmanned teaming assets across all domains' (Airbus, 2020).

Once again, one encounters the role of the network as a necessary *enabling condition* for the differentiation into 'manned' and 'unmanned'. Therefore we can draw the conclusion that the performance promises made within such AWS imaginaries are perceived in a reductionist manner if the distinction between manned and unmanned, which is undoubtedly crucial in current AWS imaginaries, is understood in such a way that AI would make a difference only on the side of the unmanned systems. In contrast, AI-based technologies through which essential tactical information and decisions are conveyed (target identification, recommendations for action and so on) are built in the operations via the networked systems. If one includes the core of the fascination of the 'network', highlighted by the imaginaries as a necessary condition, then combat clouds and related concepts are always already fully integrated into all manned operations. Hence, a central motive of so many

[11] The four companies are Hensoldt Sensors GmbH, Diehl Defence GmbH & Co. KG, ESG Elektroniksystem- und Logistik-GmbH and Rohde & Schwarz GmbH & Co.

AWS imaginaries, the loss of 'meaningful human control' over the killing decision (Bhuta et al, 2016), has to be read in a different light. The problem is *not* that autonomous machines can, in principle, kill without human control; it is that more and more human actions in the context of warfare are influenced by AI – that is, by automated and networked information systems which are potentially autonomous themselves. Hence, effects of autonomy must first be analysed at the level of networks and second at the level of effectors (for example, a drone or some other application of robotics). This leads to the conclusion that the analysis of autonomy in weapons systems and human/machine teaming should also be guided by an analysis of the imaginaries revolving around networked information systems and thus of the technological environment.

Such an argument for refocusing the debate on AWS imaginaries can be further backed up by looking at FCMS GbR's PR materials for its '4π Meta Sensor Effector System'. This system is a central component of the NGWS, with FCAS at its core. FCMS GbR state on its website:

> The NGWS team must therefore create a network of its own, equipped with appropriate capabilities and linked to the corresponding sensor and effector technology. This is the best way to ensure that opposing air defence networks can be effectively tracked, targeted and engaged. For this purpose, a large number of tasks in the fields of reconnaissance, target detection, target recognition and tracking and target engagement (kinetic and electromagnetic) as well as self-protection need to be coordinated and implemented synchronously with the highest level of precision and in a highly automated way. Humans must be integrated into the decision-making processes. (FCMS GbR, 2023)

Almost reflexively, it is mentioned at the end of the preceding quote, in the face of a clear vision of automation, that humans should still be integrated into the decision-making processes. Above all, it is interesting that the whole quote is about a vision of the automation of entire networks, which in turn are supposed to fight other networks. At this point, it becomes apparent – in contrast to older NCW concepts – that in recent imaginaries, a strong anticipation of a symmetrical threat by other highly developed foes is present. Hence, the ideas generated by the core of fascination network articulate a cybernetic vision of robustness, resilience and situatedness of one's own network, that is, the differentiation of the entire network as a temporarily independent and self-regulated – and thus presumably *autonomous* – system in itself. But there is a second side to all this. The website states:

> Future-proof solutions must meet suitable framework conditions in order to be able to provide a multidimensional '4π meta sensor-effector

> system'. In addition to a suitable basis of different multifunctional sensor and effector systems and data links, flexible interfaces to and the capabilities of mission management functions need to be taken into account. The integration into a variety of platforms as well as into the C4I environment entail additional framework conditions. (FCMS GbR, 2023)

Despite all the temporary situatedness and local specificity of the imagined combat network and its resulting effects of autonomy, the idea of developing 'frameworks' for an universal implementation of the technological vision as a whole is by no means dropped. A universal metasystem is imagined, which is able to integrate even the most heterogeneous systems. This 'universality' is, on the one hand, addressed as a technological problem, calling, for example, for 'flexible interfaces' in order to connect different technical systems, which are already networked. On the other hand, it is identified as a social problem in the sense of changing military practices such as planning or command and control:

> Accompanying to this, a suitable framework in the fields of education, mission planning, mission implementation as well as mission post-processing but also logistics needs to be established. The objective must be to implement the functionalities of networked effector and sensor systems, independent from a specific platform as well as across all performance dimensions (Air, Land, Sea) to allow the most effective use by the armed forces. (FCMS GbR, 2023)

It is obvious that the idea of network-centricity has undergone a substantial transformation in accordance with the evolution of network technology. This can be analysed as a shift on the level of network as a core of fascination for AWS imaginaries. While adapting to new technological conditions, the imaginary articulated around 2020 articulates the idea of universal connectedness through networking as well the idea of superior operational effectiveness, which was already at the heart of the NCW imaginary in the early 2000s. The emphasis on 'frameworks', which fulfil a translatory function, sheds light on the main difference between the two eras. The desired meta-network is not about the networking of different separate entities (the 2000s), but of different networks themselves (the 2020s). This is considered to be a sociotechnical issue of integrating and hierarchizing different 'networked effector and sensor systems' according to the specificity of one universal standard for network-centric operations. An illustration of these shifts is given in Table 1.

Regarding the consequences for the analysis of AWS imaginaries, we are now in the position to conclude that insofar as AI plays a decisive role for

Table 1: Network imaginaries 2000/2020

Network imaginaries around 2000	Network imaginaries around 2020
Internet 1.0 (1990s) as a reference	Internet of Things (IoT) as a reference
Network as a metaphor for society	Ubiquitous computing as societal reality
Network as condition for human decisions	Machine autonomy as a network-effect
Human information superiority as a goal	AI-based information as operational basis
Synchronization of different entities	Manned-unmanned teaming
Network as a goal (system of systems)	Networks are autonomous entities
Networking of military entities	Networks as entities for warfighting
Networking effectors and sensors as a goal	Use of sensor-effector-networks
Interfacing between different actors	Interfacing between complex networks
Enabling accelerated human decision making	Using human skills in fighting networks

information processing in the network as well as for the performance of the network as such, an infrastructural concept of autonomy and autonomous systems is emerging in the debate on AWS that cannot be understood along the lines of the distinction 'manned = human' vs. 'unmanned = machine (AI)'. What has to be understood is that in current AWS imaginaries, the *autonomy of AWS is a network effect*. Moreover, the *universal applicability* of temporally situated, robust combat networks which (under threat) can act as autonomous networks is the whole idea. Being a crucial part of AWS imaginaries, NCW today consists of the idea of building temporary, AI-enabled fast-acting, adaptive and situationally adjusted autonomous combat networks, which can be attached to a higher-level network at any time and are not fixed to a specific domain of warfare, and that consist of manned *and* unmanned elements.

Conclusion

Obviously, there is a close connection between AWS imaginaries and AI in public perception. While this might be helpful in informing the public that AI plays a significant role in warfare, it distorts the analysis of the infrastructural normalization of complex network-centric warfare as a condition for autonomy. As a core of fascination, the 'network' continues to be an attractor for ideas, visions and promises within the military and the arms industry. Given this premise, a broader analysis of AWS imaginaries in various public spheres might articulate the assumption that the more general the addressed public is, the more imaginaries are associated with traditional stereotypes from the tradition of AI narratives. But this analysis would not be complex enough. A brief glimpse into science fiction is helpful here.

Within the debate of AWS imaginaries, the sometimes grossly misinterpreted *Terminator* franchise has always revolved around the idea of a network-based AI. For example, *Terminator 3: Rise of the Machines* (2003) relied on analogies to the IoT such as the weaponization of the material environment like trucks and other machinery. Another example would be the critically acclaimed and aforementioned series *Westworld*. The action scenes of the third season, which was released in 2020, provided unfiltered scenarios of network-centric war using drones and other AI-enabled weapons in urban settings.

Yet, despite this visibility of the network as a core of fascination in popular culture, the question arises as to how this can be captured analytically. A helpful idea is to take the network as an attractor for 'media-centric' ideas, visions and promises. NCW imaginaries around 2000 have to play a particularly important role in this context. NCW was not only a contemporary reaction to the theory of a so-called 'revolution in military affairs' which has been in circulation at least since the 1991 Gulf War; rather, the specific revolution of this era was a *media revolution* that continues to this day. Talking about the network as a core of fascination, Glaubitz et al (2011: 123, author's translation) state: 'Around 2000, knowledge about the technical, social, economic and political preconditions as well as the consequences of computer networking reached its maximum level of recognition.' Back in the day, new forms of networked, computer-based information processing and the corresponding media technologies were regarded as the technological revolution, which had to be analysed and understood. Today, the recognition of AI dominates public discourses and has, possibly, reached its 'maximum level of recognition'. However, AI depends on infrastructural networks, in which it is integrated. This aspect should not be overlooked in current AWS imaginaries precisely because the imaginary of an autonomously fighting network of material entities has taken shape across all domains of warfare. By analysing AWS imaginaries, we can see how the idea of network-centric warfare has evolved into the idea of autonomous fighting networks.

Acknowledgement

This text was written within the framework of the research network 'Meaningful Human Control – Autonomous Weapon Systems between Regulation and Reflection', funded by the German Federal Ministry of Education and Research (BMBF) (grant number 01UG2206C).

Filmography

The Matrix (1999) Directed by the Wachowskis. US: Warner Brothers Pictures.

Terminator 3: Rise of the Machines (2003) Directed by Jonathan Mostow. US. Germany and UK: Warner Brothers Pictures.

Westworld, Seasons 1–4 (2016–2022) Created by Jonathan Nolan and Lisa Joy. US: HBO Entertainment.

References

Airbus (2020) 'Multi-domain combat cloud. Collaborative combat across all domains', *Airbus*. Available from: https://www.airbus.com/sites/g/files/jlcbta136/files/2021-09/Flyer_EN.pdf [Accessed 1 September 2023].

Alberts, D.S., Garstka, J.J. and Stein, F.P. (2000) *Network Centric Warfare: Developing and Leveraging Information Superiority*, 2nd edn. Washington, DC: CCRP Publication Series.

Alberts, D.S., Gartska, J.J., Hayes, R.E. and Signori, D.A. (2001) *Understanding Information Age Warfare*, Washington, DC: CCRP Publication Series.

Anders, G. (1994 [1956]) *Die Antiquiertheit des Menschen, Vol. 1: Über die Seele im Zeitalter der zweiten industriellen Revolution*, 7th reprint, Munich: C.H. Beck.

Arquilla, J. and Ronfeldt, D. (2000) *Swarming & the Future of Conflict*, Santa Monica, CA: RAND Cooperation.

Bhuta, N., Beck, S., Geiß, R., Liu, H.-Y. and Kreß, C. (2016) *Autonomous Weapons Systems: Law, Ethics, Policy*, Cambridge: Cambridge University Press.

Biltgen, P. and Ryan, S. (2016) *Activity-Based Intelligence: Principles and Applications*, Norwood: Artech House.

Castells, M. (1996) *The Information Age: Economy, Society, and Culture, Volume 1: The Rise of the Network Society*, Oxford: Blackwell.

Castoriadis, C. (1987) *The Imaginary Institution of Society*, Cambridge: Polity Press.

Cave, S. and Dihal, K. (2019) 'Hopes and fears for intelligent machines in fiction and reality', *Nature Machine Intelligence*, 1(2): 74–78. https://doi.org/10.1038/s42256-019-0020-9

Cave, S., Dihal, K. and Dillon, S. (2020) *AI Narratives: A History of Imaginative Thinking about Intelligent Machines*, Oxford: Oxford University Press.

Cebrowski, A.K. (2005) 'The implementation of network-centric warfare', *US Department of Defense, Office of Force Transformation*, 5 January. Available from: https://apps.dtic.mil/sti/pdfs/ADA446831.pdf [Accessed 1 September 2023].

Corn, J.J. and Horrigan, B. (1984) *Yesterday's Tomorrows: Past Visions of the American Future*, New York: Summit Books.

Department of Defense (2003) *Network-Centric Warfare: Creating a Decisive Warfighting Advantage*, Arlington: US Department of Defense, Office of Force Transformation.

De Landa, M. (1991) *War in the Age of Intelligent Machines*, New York: Zone Books.

Dihal, K. (2020) 'Enslaved minds: artificial intelligence, slavery, and revolt', in S. Cave, K. Dihal and S. Dillon (eds) *AI Narratives: A History of Imaginative Thinking about Intelligent Machines*, Oxford: Oxford University Press, pp 189–212.

Eckel, J., Ernst, C. and Schröter, J. (eds) (2023) 'Tech Demo', *Navigationen – Zeitschrift für Medien- und Kulturwissenschaften*, 23, 1. Siegen University. http://dx.doi.org/10.25819/ubsi/10297

Ernst, C. (2018) 'Vernetzte Lagebilder und geteiltes Situationsbewusstsein – Medialität, Kooperation und Raumkonstruktion im Paradigma des Network-Centric Warfare', in L. Nowak (ed) *Medien, Krieg, Raum*, Paderborn: Wilhelm Fink, pp 417–449.

Ernst, C. and Schröter, J. (2021) *Media Futures: Theory and Aesthetics*, London: Palgrave Macmillan.

FCMS GbR (2023) 'Future combat mission system. Enabling air power in networked operations'. Available from: https://fcms-germany.net/en [Accessed 1 September 2023].

Flichy, P. (2007) *The Internet Imaginaire*, Cambridge, MA: MIT Press.

Freedman, L. (2018) *The Future of War: A History*, New York: PublicAffairs.

Future of Life Institute (2017) 'Slaughterbots', 13 November. Available from: https://futureoflife.org/video/slaughterbots/ [Accessed 1 September 2023].

Future of Life Institute (2021) 'Slaughterbots: if human: kill()', 30 November. Available from: https://futureoflife.org/video/slaughterbots-if-human-kill/ [Accessed 1 September 2023].

Gannon, C.E. (2003) *Rumors of War and Infernal Machines: Technomilitary Agenda-Setting in American and British Speculative Fiction*, Liverpool: Liverpool University Press.

Gershenfeld, N., Krikorian, R. and Cohen, D. (2004) 'The Internet of Things', *Scientific American*, 291(4): 76–81.

Gieryn, T.F. (1983) 'Boundary-work and the demarcation of science from non-science: strains and interests in professional ideologies of scientists', *American Sociological Review*, 48(6): 781–795.

Glaubitz, N., Groscurth, H., Hoffmann, K., Schäfer, J., Schröter, J., Schwering, G. and Venus, J. (2011). *Eine Theorie der Medienumbrüche 1900/2000*. Available from: https://dspace.ub.uni-siegen.de/handle/ubsi/567 [Accessed 1 September 2023].

Gray, C.H. (1997) *Postmodern War: The New Politics of Conflict*, New York: Routledge.

Grunwald, A. (2012) *Technikzukünfte als Medium von Zukunftsdebatten und Technikgestaltung*, Karlsruhe: Karlsruhe Institute of Technology.

Guha, M. (2011) *Reimagining War in the 21st Century: From Clausewitz to Network-Centric Warfare*, New York: Routledge.

Hellige, H.D. (1992) *Militärische Einflüsse auf Leitbilder, Lösungsmuster und Entwicklungsrichtungen der Computerkommunikation* (artec-Paper Nr. 15, März 1992). Available from: https://media.suub.uni-bremen.de/handle/elib/4688 [Accessed 7 January 2025].

Jasanoff, S. (2015) 'Future imperfect: science, technology, and the imagination of modernity', in S. Jasanoff and S.-H. Kim (eds) *Dreamscapes of Modernity. Sociotechnical Imaginaries and the Fabrication of Power*, Chicago: University of Chicago Press, pp 1–33.

Kirby, D.A. (2010) 'The future is now: diegetic prototypes and the role of popular films in generating real-world technological development', *Social Studies of Science*, 40(1): 41–70.

Lenoir, T. and Caldwell, L. (2018) *The Military-Entertainment Complex*, Cambridge, MA: Harvard University Press.

McNeil, M., Arribas-Ayllon, M., Haran, J., Mackenzie, A. and Tutton, R. (2017) 'Conceptualizing imaginaries of science, technology, and society', in U. Felt, R. Fouché, C.A. Miller and L. Smith-Doerr (eds) *The Handbook of Science and Technology Studies*, Cambridge, MA: MIT Press, pp 345–363.

Natale, S. and Balbi, G. (2014) 'Media and the imaginary in history', *Media History*, 20(2): 203–218.

Neuroth, O. (2023) 'Spanien steigt bei Rüstungsprojekt FCAS ein', *Tagesschau*, 28 April. Available from: https://www.tagesschau.de/inland/innenpolitik/fcas-kampfflugzeug-100.html [Accessed 1 September 2023].

Rosental, C. (2021) *The Demonstration Society*, Cambridge, MA: MIT Press.

Russell, S. (2019) *Human Compatible: AI and the Problem of Control*, London: Penguin.

Schröter, J., Ernst, C. and Warnke, M. (2022) 'Quantum computing and the analog/digital distinction', *Grey Room*, 86: 28–49.

Singer, P.W. (2010) *Wired for War: The Robotics Revolution and Conflict in the Twenty-First Century*, New York: Penguin.

Smith, W. (2009) 'Theatre of use: a frame analysis of information technology demonstrations', *Social Studies of Science*, 39(3): 449–480.

Sprenger, F. (2021) *Autonome Autos. Medien- und kulturwissenschaftliche Perspektiven auf die Zukunft der Mobilität*, Bielefeld: transcript.

Sprenger, F. and Engemann, C. (2015) *Internet der Dinge. Über smarte Objekte, intelligente Umgebungen und die technische Durchdringung der Wel*t, Bielefeld: transcript.

Taylor, C. (2004) *Modern Social Imaginaries*, Durham, NC: Duke University Press.

US Navy (2017) 'Perdix swarm demo', *YouTube*. Available from: https://www.youtube.com/watch?v=bsKbGc9TUHc [Accessed 1 September 2023].

Weber, J. (2021) 'Artificial intelligence and the sociotechnical imaginary: on Skynet, self-healing swarms and slaughterbots', in K. Maurer and A. Immanuel Graae (eds) *Drone Imaginaries and the Power of Vision*, Manchester: Manchester University Press, pp 167–179.

Weiser, M. (1991) 'The computer for the 21st century', *Scientific American*, 265(3): 94–104.

Wolski, R. (2022). 'Aim high!', *Behörden Spiegel. Moderne Streitkräfte. Kampfflugzeuge der sechsten Generation, Sonderpublikation*: 4–5.

Governing Autonomies: Imagining Responsible Artificial Intelligence in the 'Future Combat Air System' European Armament Project

Jens Hälterlein

Introduction

This chapter addresses the rise of responsible artificial intelligence (RAI) as a mode of governing autonomous weapon systems (AWS). The ambiguity of the title 'governing autonomies' is deliberate: on the one hand, RAI aims to ensure human control over AWS – that is, the governing of machine autonomy. On the other hand, RAI aims to subordinate AWS to the sovereign decisions of humans – that is, AWS are governed by human autonomy.

These power relations between humans and machines are the matter of concern in the controversy on AWS. The development and (potential) use of AWS is criticized by various actors from politics, academia and civil society, who caution against the loss of human responsibility and accountability, and call for a meaningful human control to be implemented or a ban on these weapons to be enforced. While these criticisms have not yet led to a legally binding regulation at the international level, it can be seen as their unintended consequence that in military strategy, voluntary commitments to the guiding vision of a responsible (use of) AI have been expressed (French Ministry of Armed Forces, 2019; US Department of Defense, 2020; NATO, 2021, Government of the Netherlands, 2023). What the operationalization of this guiding vision of can look like on the level of concrete technology development can be observed with regard to the trinational project *Future Combat Air System* (FCAS). As of 2023, FCAS is a cooperation of Germany, France and Spain, while Belgium has observer status and may become a fully fledged project partner at a later stage. The project is Europe's most ambitious armament project (Karakas, 2021). It is envisioned to become the core of the European air forces from around

2040. This chapter presents an analysis of how a responsible use of AWS is imagined within the working group *Responsible Technology for an FCAS*, a forum that has been established by two German key actors for the purpose of handling ethical and legal issues of FCAS. To date, the forum is the only institutionalized form of ethical reflection within the project. My focus on German actors and their statements is therefore a necessary result of the structure of the FCAS project.

The focus on imaginaries is inevitable in this case, as FCAS is a future vision and technological innovation is still at the level of developing prototypes. However, shedding light on a project at an early stage implies that there is still the possibility to intervene before technologies are used in practice, which in the case of FCAS means AWS that are used in military operations. This intervention is all the more important, as the way in which a responsible use of AWS is imagined in FCAS falls short. As I will argue in the following, this failure is closely related to the enactment of a neoliberal rationality of individual responsibilization and indirect behaviour management.

By 'neoliberal rationality', I refer to a specific historical form of the government of self and others. Neoliberal government aims to establish economic rationality across society (Foucault, 2008; Bröckling et al, 2010; Lemke, 2021) and is enacted via specific modes of subjectification (Burchell, Gordon and Miller, 1991; Dean, 1999). It favours indirect or remote forms of control and thus operates through 'chains of enrolment, "responsibilization" and "empowerment" by individuals who adopt a new and specific mode of governing the self' (Barry, Osborne and Rose, 1996: 12). At the centre of this neoliberal rationality is the interpellation of the entrepreneurial self to permanently adapt and self-optimize in view of ever-changing environmental conditions (Bröckling, 2016). It aims at bringing about a prudent and active self-conduct of subjects by mobilising self-reflection and engagement. Thereby, state tasks and responsibilities are shifted to the individual's sphere of responsibility. In this respect, responsibilization refers to the practice of assigning individual responsibility so that an individual 'would produce the ends of government by fulfilling themselves rather than being merely obedient' (Rose, O'Malley and Valverde, 2006: 89). Instead of addressing social problems by changing the structural conditions of poverty, unemployment or crime, individuals are held responsible for their own fortune and misfortune. However, neoliberal forms of the 'conduct of conduct' (Lemke, 2001: 191) may as well seek 'to govern the "environment" of human and nonhuman entities rather than operating directly on "subjects" and "objects"' (Lemke, 2021: 168). This logic of 'environmentality' (Lemke, 2021: 169) is less concerned with targeting individual behaviour, but rather seeks to alter the material conditions and contexts of this behaviour in order to implement regulatory strategies. In situational crime prevention, for instance, opportunity structures are manipulated in such a way (more officers on patrol, more surveillance cameras, tougher penalties and so on) that it becomes increasingly irrational to commit a crime.

It is precisely this neoliberal rationality that structures many of the AI strategies that have been presented by national governments and supranational organizations over the last few years. As Hälterlein (2023; 2024) has argued, both the strategies of the European Union (EU) and of Germany aim to govern the risks of nonmilitary AI systems and to promote trust predominantly through fostering 'soft law' – voluntary codes of conduct and ethics by, in and for design. In the terminology of the EU's High-Level Expert Group on Artificial Intelligence, *ethics by design* refers to risk management through the implementation of ethical and legal principles (fairness, explainability, transparency, security and so on) from the beginning of the design process (High-Level Group on Artificial Intelligence, 2019: 21). Mandatory legal requirements such as certification and conformity assessment by public authorities should only apply to those AI systems that are classified as high-risk applications. In case of AI systems posing very high risks, restrictions of use or even a legal ban might be issued. However, AWS are not affected by these regulations. They are deliberately excluded from both AI strategies with reference to military AI strategies that have yet to be developed. Instead, as I will show in the following sections, the FCAS project and its imaginary of RAI enact a neoliberal rationality of individual responsibilization and indirect forms of behaviour control that the aforementioned AI strategies foresee for low-risk applications only. Moreover, I will demonstrate that these indirect forms of control are inadequate to secure a responsible use of AWS. This, in turn, reinforces the call for the binding legal regulation of these weapons or even their ban.

The structure of the chapter is as follows: the first section will provide an overview of the FCAS project. Subsequently, I will describe how AI is imagined to enhance human decision making and to enable individual responsibility as the means to ensure a responsible use of AWS within the *Responsible Technology for an FCAS* working group. I will then point out the liberal anthropology underlying this imaginary of RAI and show how neoliberal forms of government are brought into play in order to realize it. The last section will analyse the shortcomings of this approach by means of the so-called FCAS Ethical AI Demonstrator.

1. The future combat air system of systems

FCAS does not refer to the future vision of a single weapon system, but rather to a network of several weapon systems which consists of existing weapon systems (for instance, the *Eurofighter* combat jet and the *Tiger* combat helicopters), but also new developments such as the *Eurodrone* and, in particular, the *Next Generation Weapon System*, consisting of a new combat aircraft, the *Next Generation Fighter*, *Remote Carriers*, a technical term for drones that carry a range of payloads (both sensors and weapons), and a digital infrastructure that is to connect all elements of FCAS, the so-called *Combat Cloud* or *Multi-Domain Combat Cloud* (Bundesverband der Deutschen Luft- und Raumfahrtindustrie e.V., 2021). Other manned and unmanned components could additionally be integrated into FCAS. The expected military added value of FCAS thus lies less in the individual weapon systems or their sum, but rather in the way in which they are networked with each other in a system of systems approach. In FCAS, the integration of AI is envisioned in different ways with a focus on different elements of the system of systems. First, the *Next Generation Fighter* would operate together with the *Remote Carriers* in the form of *Manned-Unmanned Teaming*. This Manned-Unmanned Teaming will require the use of AI to enable drones to operate autonomously to a certain extent (Airbus Defence and Space, 2020a). Second, the *Combat-Cloud* is envisioned to include an algorithmic decision support system that would enable faster loops of the Observation Orientation Decision Action (OODA) cycle (Klauke, 2021). Another characteristic of the project is the institutionalization of the handling of ethical and legal issues by means of a forum established specifically for this purpose: the *Responsible Technology for an FCAS* working group ('AG Technikverantwortung'). The working group was established in 2019 by the German Fraunhofer Institute for Communication, Information Processing and Ergonomics (FKIE) and Airbus Defence and Space, and brings together authorities such as the German Federal Ministry of Defence and Federal Ministry of Foreign Affairs, but also stakeholders from academia, think tanks and church institutions (Keisinger and Koch, n.d.). Despite its claim to discuss normative issues on behalf of the entire FCAS project, the working group has so far been a purely German affair. In the following, I will exclusively refer to documents and statements linked or published on the homepage of the working group. However, individual documents and statements are not necessarily representative of the opinion of all members of the working group.

2. A European way of exercising control over autonomous weapon systems

Not surprisingly, hardly any of the documents and statements published on the homepage of the working group question the development of AWS as part of FCAS. A certain level of 'technical autonomy'[1] is considered

[1] The term 'technical autonomy' is defined as the 'ability of an artificial system to perform even highly complex tasks in an automated fashion' (Azzano et al, 2021: 4).

inevitable, since without the high degree of automation it enables, FCAS would be ineffective and thus futile in future military conflicts with faster-acting opponents and more complex situations (Koch and Keisinger, 2020). For FCAS, however, there would still be the opportunity 'to go for a European way that keeps the overall system under control of an informed, aware, and accountable human operator, which is equipped with means of control that are meaningful to the required and specified level' (Azzano et al, 2021: 9). This claim is justified by two assumptions: that AI-based situational awareness can partially alleviate the 'fog of war'; and that human autonomy can master this powerful tool.

2.1 Artificial intelligence-enabled human decision making: from meaningful human control to accountable responsibility

According to the promoters of FCAS, the AI-based processing and analysis of large amounts of data would provide a degree of situational awareness as has never been the case before. In the past, an enormous quantity of data was seen as a problem because it made military decision making more demanding and ultimately based on a high level of uncertainty in situational awareness. However, within the future vision of FCAS, data abundance is exactly what is desired, since it would make decision making easier and hence enable 'decision superiority' (Airbus Defence and Space, 2020b). As the use of AI in FCAS unburdens human operators from routine or mass tasks, they would remain capable of 'appropriately acting even on short time scales in the complex "technosphere" of modern warfare with spatially distributed and highly agile assets', as Wolfgang Koch, Professor for Computer Science at University of Bonn and lead scientist at FKIE, states (Koch, 2022: 1). In light of this technoscientific promise, the 'fog of war' described by Clausewitz would not disappear completely, but it would be less heavy. In the words of Colonel Hubert Saur, a retired German combat pilot who now works for Airbus, 'AI and Human-Machine Collaboration will contribute to information and decision superiority ensuring meaningful control throughout the mission cycle' (2021). Thus, the incorporation of AI into FCAS would not lead to a loss of meaningful human control, but to improved controllability of the situation and to 'accountable responsibility' (Koch, 2022: 4) – that is, 'the overall system design must guarantee that always a distinct 'somebody' is responsible' (Koch, 2022: 4).

The introduction of the notion of 'accountable responsibility' by Koch is meant to broaden – or rather replace – meaningful human control as a regulatory ideal as the former is elevated to the normative standard or touchstone for the FCAS project (Koch, 2022: 4). On the one hand, it would be more fundamental than meaningful human control, which is, in this case, understood as the inclusion of the pilot in all processes (that is, the human-in-the-loop) or the permanent tracking of all processes by the pilot (that is, the human-on-the-loop). On the other hand, unlike meaningful human control, it would be possible to achieve accountable responsibility not in spite a high level of automatization, but because of it. In sum, the introduction of the notion of 'accountable responsibility' into the imaginary of RAI in FCAS alters the normative thrust of the critique of AWS as it equals more machine automation with more human control.

2.2 The autonomous human operator and the powerful tools

Besides AI-based situational awareness, there is a second prerequisite for accountable responsibility. This prerequisite is 'human autonomy': 'To speak of responsibility is only reasonable if it is assumed voluntarily. Responsibility, thus, presupposes the notion of a "free will" and an Image of Man [*sic*] as a free and "autonomous" person' (Koch, 2022: 4). As Koch continues to elaborate, '[t]he concepts of mind and will, and therefore of consciousness and responsibility bring natural beings into view that are "somebody" and not "something", i.e. persons, and open up ethical dimensions' (Koch, 2022: 2). Accordingly, algorithms are not to be seen as endowed with reason and agency, but as 'cognitive and volitive tools' (Koch, 2022: 4) whose only function is to assist the intelligent minds and autonomous wills of human beings. What AI would enable is hence not autonomous machines, but machines 'that greatly enhance the perceptive mind and the active will of persons, who alone are capable to perceive intelligently and to act autonomously in a proper sense' (Koch, 2022: 1). Koch differentiates two types of these human enhancement machines: *cognitive machines* that process data and create situational pictures, which in turn are used by human beings for intelligent forms of understanding and planning and *volitive machines* that execute deliberately taken decisions of a responsible human endowed with a free will and transform these 'into complex command sequences to control networking platforms, multifunctional sensors, and effectors' (Koch, 2022: 2). While these *cognitive* and *volitive machines* are supposed to support human understanding of a situation and human choices, any ethically and legally acceptable use of these machines must rely on appropriate human situational awareness and appropriate normative orientation (Koch, 2022: 3). The machine simply follows orders, meaning that 'accountable responsibility' may only reside at the human while an AI-based machine remains a thing, a tool, subordinated to the will of its commander. Therefore, the new capabilities of AI-driven machines require 'naturally intelligent critical capabilities of military decision makers toward AI' (Koch, 2021a: 105). Otherwise, there would be the danger of obedience to the output of a system, and the danger of refusal to bear responsibility (Koch, 2021a: 105).

3. Producing the subject of accountable responsibility

Koch's 'anthropocentric' account of AI is based not only on a traditional understanding of technology but also on a liberal understanding of the subject. This subject is endowed with natural rights and freedoms, is rational and possesses free will. Yet, from a perspective of Foucauldian discourse analysis, this liberal subject is rather to be understood as a historically specific formation of power/knowledge, a product of the discursive construction of a historical truth that turns human beings into liberal subjects. However, at the same time, these 'natural qualities' of the liberal subject have to be cultivated; they are a potential inherent to humans that must first be fully exploited – otherwise, certain individuals or subpopulations would demonstrate a lack of rationality, prudentialism and responsibility. Therefore, the guiding principle of (neo)liberalism to acknowledge and praise individual freedom corresponds to a governing of the way in which freedom is exercised. Interventions into individual freedom become the necessary downside of this freedom as interventions are employed to produce appropriate forms of self-conduct. These interventions can take the form of educational institutions, counselling services, psychotherapy or nudging to foster healthy consumer behaviour.

But how can the required critical capacity on the part of the human operators of the FCAS be achieved? How can these operators become the responsible subjects they are supposed to be and the regulatory ideal of 'accountable responsibility' become a reality? From Koch's perspective, the answer is twofold. First, digital ethos and morality need to be cultivated in parallel to technical development. Military training and personality development should focus on ethical capacities and leadership at all levels of application of AI-based machines (Koch, 2021a: 102). Second, AI-based machines must be designed in such a way that they train the 'vigilance of their users and convey to them how the machine solutions were created' (Koch, 2021b: 7). As we will see, these two strategies correspond to topical modes of the neoliberal conduct of conduct, through which government can be exercised: the responsibilization of the individual and environmentality.

3.1 Innere Führung*: making the responsible operator*

Given the need for critical capacity at the part of the human operator, *Innere Führung* (leadership development and civic education) (Bundeswehr, n.d.b) is proclaimed by Koch as the guiding principle for any responsible use of AI-based machines in FCAS. Since the foundation of the German Army in the 1950s, the *Innere Führung* concept is the underlying philosophy of leadership valid for the German soldiers. Although the concept has a specific national and historical context of origin and has remained a uniquely German phenomenon, it is elevated by Koch to become the universal guiding principle for RAI in the military domain. *Innere Führung* is the normative ideal of the commitment of the soldier to liberal moral-ethical standards and responsibility for the protection of the German democratic constitutional state. Together with the command and control principle *Auftragstaktik* (mission command/mission-type tactics) (Geyer, 2018), it is seen as one of the two cornerstones of today's German Army. Since in *Auftragstaktik*, subordinates need to pursue the ends of a mission independently, meaning that they are largely free to choose a way to fulfil these ends responsibility, good judgement and decisiveness are required of every soldier (and not only of military leaders). Only on the condition that soldiers are intrinsically motivated and consciously responsible for their own actions and that they strive to pursue ends while at the same time recognizing the limits of acting and obeying, a leader can trust his subordinates (Führungsakademie der Bundeswehr, 2014: 6).

Hence, *Innere Führung* can be seen as the predominant form of subjectivation of the German soldier: it provides the blueprint for the military individual to become the subject she or he is supposed to be. Like any, *Innere Führung* must be actively appropriated and applied. And just like any other technology of the self, it is the condition of the governing of others – in this case, the subordinate soldiers. However, in the case of *Innere Führung*, the responsibilization of the individual not only addresses their self-conduct and their conduct of other humans but also their capabilities and duties regarding a responsible government of things.

Since its introduction as a leadership concept, *Innere Führung* was also meant to deliver the means to grapple with the far-reaching consequences of decisions, given the enormous destructiveness and range of modern weaponry. As Wolf von Baudissin – a German general who developed the guiding visions for the German Army after the Second World War and is considered one of the 'fathers' of the *Innere Führung* concept – noted, the most highly technical nature of combat requires that responsibility is seen and borne at many lower levels (Baudissin, 1969: 234). Therefore, he continues, the training of soldiers must aim to challenge their responsibility and make them experience the consequences of their actions and omissions. While Baudissin's argument referred to the Cold War era and its weapons of mass destruction, it is no less valid concerning the (hot) wars of the future and its AI-based weapons systems, at least according to Koch. The guiding principle of *Innere Führung* must therefore be transferred to their application. It is this technology of the self that would equip humans with the 'digital ethos' (Koch, 2022: 3) that provides the means to responsibly use AI-based weapon systems that threaten to escape (meaningful) human control.

3.2 Ethics and reflectivity by design: the environment of responsibility

But how can a high level of automatization, as envisioned in FCAS, be aligned with the *Innere Führung* of their human operator, as envisioned by Koch (2022)? In principle, 'accountable responsibility' would exist as long as automated decisions are only permitted on the basis of a normative set of rules defined by humans in advance and insofar as human operators of an automated system can assess whether a set of rules defined and implemented in advance is applicable in a given situation. In the first place, applicable ethical and legal norms have to be operationalized, translated into technical standards and implemented into design choices (Koch and Keisinger, 2020). Accordingly, *ethics-by-design* is highlighted as a fundament of RAI in FCAS. However, Koch considers it equally important that the comprehensibility, plausibility or explainability of AI-based results is guaranteed. Only in this way are humans enabled to consciously weigh up recommendations, instead of confirming them simply based on trust (Koch, 2022: 5). As he states: 'it must be clarified on which technically realisable basis a human operator can ultimately make balanced, consciously considered decisions regarding the use of armed force (meaningful authorisation)' (Koch, 2021b: 5). This approach to enable RAI in FCAS by aligning the system design with given normative criteria is consistent with the operational mode of environmentality – the altering of the material conditions and contexts of behaviour in order to indirectly influence it. Regarding FCAS, the aim is to alter the system design in such a way that it becomes likely that the regulatory ideal of 'accountable responsibility' becomes a reality.

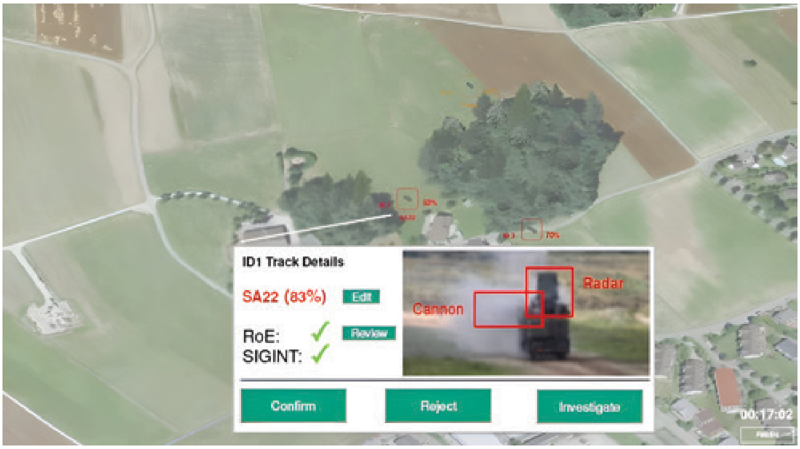
To date, the most elaborate contribution of the working group regarding this aim is a White Paper entitled 'The responsible use of artificial intelligence in FCAS – an initial assessment' (Azzano et al, 2021), which was written by an independent group of Airbus Defence and Space engineers. Concerning the international debate on AWS and meaningful human control, the authors state that there is a trade-off between human intervention and operational efficiency (Azzano et al, 2021: 3). In the following chapter, I will show that this problem remains unresolved within the project. The focus of the White Paper is then on the work of the *High-Level Expert Group on AI* (AI HLEG) set up by the European Commission. In its initial report, the AI HLEG states that a *trustworthy AI* must be lawful, ethical and robust (High-Level Group on Artificial Intelligence, 2019). This report was complemented by another document called the 'Final assessment list for trustworthy artificial intelligence', which lists 131 questions proposed as a practical tool to check the trustworthiness of a particular AI system under development – the so-called Assessment List for Trustworthy AI (ALTAI) (High-Level Group on Artificial Intelligence, 2020). After having applied the ALTAI to a number of relevant use cases, the authors of the White Paper state that there is a 'need to both define and integrate Meaningful Human Control, either in-the-loop or on-the-loop, at cardinal design points within FCAS' (Azzano et al, 2021: 2). Moreover, this will have to be done in a flexible manner that allows the tailoring of meaningful human control according to the type of mission and the applicable rules of engagement (RoE) – the directives to military forces that define the lawfulness of performing an attack (Azzano et al, 2021: 4). Hence, 'human supervision and intervention will be required through the employment of a clear, unambiguous and coherent concept for human accountability, which enables the user to understand the level of accountability and ethical responsibility for every decision that is made' (Azzano et al, 2021: 7).

4. Putting autonomy to the test: the Ethical Artificial Intelligence Demonstrator and its shortcomings

One of the use cases discussed in the White Paper is the use of AI to detect and identify potential targets as part of the Targeting Cycle, the sequence of the operational activities Find/Fix/Track/Target/Engage/Assess (Azzano et al, 2021: 7). This use case has also been scrutinized by the FCAS's 'ethical demonstrator initiative'. The *Ethical AI Demonstrator* (E-AID) simulates the application of AI in FCAS on the basis of the prototype of a decision support system. Simulations are run on scenarios closely coordinated with the German Army. The aim is to gain a realistic picture of the possibilities, limits and implications of AI in defence by means of concrete examples and to provide a first 'hands-on' step towards an 'ethics-by-design' methodology which then can be integrated into an overall FCAS design process (FCAS Forum, 2021). As a testing device, the E-AID is expected to clarify which system design is best to deliver 'reflective assistance' to human operators – that is, to enable them 'to make balanced and conscious decisions regarding the use of weaponry based on artificially intelligent automation' (Koch, 2022: 5). Moreover, given the need to cultivate *Innere Führung*, the 'E-AID may serve as a simulator for training the responsible execution of the targeting cycles of future combat air systems such as FCAS' (Koch, 2022: 7). In terms of soldierly subjectivation, the E-AID is expected to deliver an experience of individual responsibility and to foster *Innere Führung* in the face of the consequences of one's own decisions in a highly mechanised environment.

Within the E-AID scenario *Find Fix Track Application with AI for Automated Target Recognition*, the mission is the elimination of enemy air defence using remote carriers equipped with sensors that collect data on positions of military equipment supporting the enemy air defence. The output of the automated target recognition is shown on a graphic interface highlighting relevant objects and providing basic context information (for example, type of detected vehicle or certainty level). The guiding question in this

Figure 1: Interface of the E-AID



Source: FCAS Forum (2021)

scenario is how tasks can be delegated to an AI in an accelerated decision cycle without violating applicable military RoE.

Figure 1 is a screenshot of the interface of the E-AID taken from a presentation given during a meeting of the working group (FCAS Forum, 2021). The aerial view of the site shows that several objects have been identified and each of these objects has been assigned an identifier, a label and a probability value. In a smaller browser window, the 'track details' are given for one of the identified objects (ID1). With a probability of 83 per cent, the object is said to be a Russian SA22 (Pantsir-S1) tank which is an enemy weapon system in the given scenario. On a magnified image, two details of the object are marked by red squares and are classified as 'cannon' and 'radar'. Green ticks are placed behind 'RoE' and 'SIGINT' (short for Signal Intelligence). In addition, several fields can be clicked on by the user, such as 'Edit', 'Review' and 'Investigate'. According to Koch, this design will enable the human operator 'to confirm the target by visual address and to understand in the magnified section by means of appropriate highlighting of Explainable AI (XAI) which has recognized elements of the tracked object' (Koch, 2022: 5).

In the following, I will show that this promise cannot be fulfilled and that 'accountable responsibility' is not given. The possibility for the user to obtain background information on the individual outputs of the system is an essential element of the ethics-by-design approach and should guarantee both the 'trustworthiness' of the AI-generated output and the 'accountable responsibility' of the human decisions. However, if the user does all (or at least many) of the possible background checks on all (or at least many) of the identified objects (by clicking on the 'Edit', 'Review' and 'Investigate' buttons), the speed advantages of automating the OODA cycle are lost. If, on the other hand, the user abandons these options, she or he simply has to trust the system. This would be even more problematic, insofar as it would imply a leap of faith which the system cannot justify at all. There is no reason to assume that a FCAS is less prone to errors in data analysis than other AWS. Hence, the only difference is that the human operator should be given the opportunity not to convert these erroneous recommendations into deeds which would only be possible at the expense of the lethal performance of the overall system. Thus, despite the ethics-by-design approach, there would be a trade-off between 'accountable responsibility' and the efficiency of the AWS.

Yet, even more fundamental is the question of the extent to which these background checks would actually enable a human operator to understand the accomplishment of the output. Or to put it another way: what does it mean to understand or to explain the output of an algorithmic decision support system? The answer to this question is not straightforward, and it is certainly not in the FCAS scenario at hand. Is it sufficient to highlight the radar and the gun in the magnified section as an explanation for the categorization as a SA22? How then is the human operator to understand the conclusion of the system that the details shown in the image are a radar and a gun or – even more complicated – that the radar and the gun belong to a SA22 and not to another military vehicle? A human understanding of this output would ultimately require an explanation of the computational methods and the data involved. However, whether this is 'understandable' depends on the complexity of the algorithms and the expertise of the human operator. With regard to machine learning algorithms (which are likely to play a decisive role), even providing open-source code and training data might not be sufficient to make an output fully understandable, especially not for end users from the military who are normally laypersons in computational science. Moreover, if artificial neural networks are used, which is often the case in computer vision applications such as target recognition, even experts in the field are unable to understand the operations of the system in greater detail. Hence, an AI-based system like the one presented in the E-AID will cause accountability problems even when full transparency is given (Hälterlein, 2021).

Another crucial problem is the way the algorithmic decision support system reduces complexity. The situation shown in Figure 1 is constructed in such a way that there are only two vehicles visible at all, both of which are classified as 'targets' with a relatively high probability (83 per cent and 79 per cent). If the targets are not military vehicles but human combatants, it is hard to imagine how a trustworthy output can be generated without a deeper understanding of the situation and, even more so, how the output can be made understandable for a human operator who is to be held accountable. But even an urban situation or a situation with a busy street in which numerous vehicles can be seen, which could be both civilian and military and, in the latter case, could be part of one's own troops as well as part of the enemy's troops presents a completely different challenge. In this case, the question arises as to whether all vehicles should be automatically classified and presented to the human operator accordingly, or whether an automatic pre-selection should be carried out. Above a certain number of vehicles and with regard to the need to reduce complexity and the hoped-for speed advantage in the OODA cycle, the situation would only be manageable through pre-selection. However, this would result in a follow-up problem: to generate a number of possible targets that is manageable for the human operator, an arbitrary decision threshold of probability must be set that separates 'likely targets' from 'unlikely targets' of which only those above the threshold (the 'likely' ones) are highlighted. However, this choice is not only arbitrary but also has an impact on the accuracy of the system: if the threshold is set low, more objects that are actually nontargets will be classified as 'likely targets' (false positives); if the threshold is set high, more true targets will not be highlighted by the system (false negatives). Depending on the context or use case, system designers or operators will choose a threshold that they deem appropriate.

In the next step, the output of the system has to be evaluated. At this point, the setting of the threshold affects human judgement as well. Empirical research on the use of algorithmic decision-making (ADM) systems (Skitka et al, 1999) found that users hardly question the output of the systems and even tend to regard it as infallible – the so-called 'automation bias'. According to Cummings (2015), automation bias effects in interaction with automated decision support systems have contributed to several fatal decisions, including the US Army's Patriot missile system shooting down a British Tornado and an American F/A-18, killing three aircrew in 2004 during the Iraq War. According to Parasuraman and Manzey (2010), automation bias depends, among other factors, on the level of automation, cannot be prevented by training or explicit instructions to verify the recommendations given by a system, and can affect the decision making of individuals as well as in teams. In the case of FCAS, metrical probability calculations (83 per cent) might even reinforce this effect.[2] An automation bias often leads to two types of errors: in a 'commission error', users follow an erroneous recommendation of an ADM system. Applied to the given scenario, this would mean that they

[2] The probability of the correctness of the classification 'target' expressed in a numerical value suggests that this is a mathematical calculation, the result of which is factually correct and thus represents an objective basis for decision making for the human operator. However, the calculated metric value can be wrong just like a 'simple' nominal distinction between target and nontarget. In this respect, it is only a matter of pseudo-objectivity.

regard false positives as true positives and take deadly actions accordingly. In the case of an 'omission error', users overlook critical situations if these are not identified by the system. Applied to the given scenario, this would mean that they would not take notice of a 'true' target, but only concentrate on the 'likely' targets highlighted on the interface – which might have deadly consequences as well. Depending on the setting of the threshold, either omission or commission errors are more likely to occur. Hence, it is the interplay of humans and algorithms that produces 'right' or 'wrong' decisions in a complex way. In light of this, it seems inappropriate to hold the human operator fully accountable for the consequences of these potentially lethal errors. This also shows the inadequacy of neoliberal logics of governing the risks of military AI. Neither individual responsibilization nor environmentality is sufficient to provide for RAI in FCAS.

Conclusion

In the FCAS project, a responsible use of AWS is imagined as an 'accountable responsibility' that can be enabled by means of a 'digital ethos' on the part of the human decision maker and ethical design on the part of the ADM system. According to this imaginary, only humans can act morally based on their autonomy and take responsibility for their decisions, while machines can and must be designed in such a way that they support humans in making responsible and reflective decisions. In order to ensure that these requirements are met, this chapter has revealed that two neoliberal technologies of government are enacted: the individual responsibilization of human decision makers and the environmental structuring of the material conditions and contexts of their decisions. However, as the analysis of the E-AID has shown, neither individual responsibilization nor an ethical design is sufficient to enable 'accountable responsibility'. Moreover, considering the challenges and the functioning of an ADM system like the one simulated in the E-AID, it is highly questionable if individual responsibility and accountability – or, in Koch's terminology, 'accountable responsibility' – can be given. If at all, these can be realized in manageable and unambiguous combat situations (for which the question of the added value of an algorithmic decision support system would arise). But even then, it is questionable whether the output of a system can be explained in such a way that a human can actually understand it.

The statement, made by some members of the *Responsible Technology for FCAS* working group, that the term 'autonomous weapon systems' is misleading because it implies that machines define and attack targets without any human intervention (Bossong et al, 2022) can be agreed with. After all, humans are involved in these processes on a wide variety of levels – long before a final button is pushed or not. However, the notion of 'human autonomy' underlying the concept of 'accountable responsibility' is equally misleading, since it assumes that a human being can be(come) the master of these highly automated processes and make sovereign decisions. After all, machines and algorithms are involved in these decisions on a wide variety of levels. Therefore, exercising and enabling the responsible use of AWS represent impossible tasks for the actors addressed by this neoliberal interpellation.[3] Thus, since the E-AID precisely leads to the experience of the impossibility of individual responsibility and accountability in the context of the use of AWS, it provokes several questions: would *Innere Führung* in the case of a FCAS not rather imply that a soldier categorically refuses to take a 'decision' based on an ADM system, be it by pressing the 'Confirm' button or the 'Reject' button? And should value-based engineering not rather mean refusing to develop military technology at all, be it with or without an ethics-by-design approach? And finally, should the principle of 'accountable responsibility' not be understood in a completely different way, namely as the decision to oppose the development and use of AWS? However, whether this 'digital ethos' of 'I would prefer not to' would be sufficient to stop the development and use of AWS on a global scale is doubtful.

Given the failure of the expert negotiations under the United Nations' Convention on Certain Conventional Weapons, I would like to end this chapter by pointing to the danger of a turning away from the possibility of a governing of AWS through hard laws and a turn towards ultimately ineffective forms of soft law such as ethics-by-design and voluntary self-commitments of states and militaries. Future approaches to regulating AI in the military should not reproduce the neoliberal dogma of replacing explicit prohibitions and control with individual responsibility and ethical framing.

References

Airbus Defence and Space (2020a) 'Manned-unmanned teaming and remote Carriers: transcending individual assets' capabilities', 8 October. Available from: https://www.airbus.com/en/newsroom/stories/2020-10-manned-unmanned-teaming-and-remote-carriers-transcending-individual-assets [Accessed 4 January 2024].

Airbus Defence and Space (2020b) 'Multi-Domain Combat Cloud datasheet'. Available from: https://www.airbus.com/en/products-services/defence/multi-domain-superiority/multi-domain-combat-cloud [Accessed 4 January 2024].

[3] The term 'interpellation' refers to the ideological conditioning of the subject by institutions and societal discourses. The interpellation and its appropriation by the addressee are seen as constitutive for subjectification.

Azzano, M., Boria, S., Brunessaux S. et al. (2021) 'The responsible use of artificial intelligence in FCAS: an initial assessment', *White Paper*, 14. Available from: https://www.fcas-forum.eu/en/articles/responsible-use-of-artificial-intelligence-in-fcas/ [Accessed 6 January 2025].

Baudissin W. (1969) *Soldat für den Frieden: Entwürfe für eine zeitgemäße Bundeswehr*, Munich: R. Piper & Co.

Barry, A., Osborne, T. and Rose, N. S. (1996) *Foucault and Political Reason: Liberalism, Neo-liberalism and Rationalities of Government*, London: Routledge.

Bossong, N., Rieks, A. and Koch, W. (2022) 'Künstliche Intelligenz für die Landesverteidigung', *Frankfurter Allgemeine Zeitung*, 2 February.

Bundesverband der Deutschen Luft- und Raumfahrtindustrie e.V. (2021) 'Das Future Combat Air System: Übersicht'. Available from: https://www.bdli.de/publikationen/fcas-i-uebersicht [Accessed 4 January 2024].

Bundeswehr (n.d.a) 'FCAS – Future Combat Air System', *Bundeswehr*. Available from: https://www.bundeswehr.de/de/organisation/luftwaffe/aktuelles/fcas-future-combat-air-system [Accessed 4 January 2024].

Bundeswehr. (n.d.b) 'Terms: The Bundeswehr from A to Z'. Available from: https://www.bundeswehr.de/en/about-bundeswehr/terms-bundeswehr [Accessed 4 January 2024].

Burchell, G., Gordon, C. and Miller, P. (1991) *The Foucault Effect: Studies in Governmentality. With Two Lectures by and an Interview with Michel Foucault*, Chicago: University of Chicago Press.

Bröckling, U. (2016) *The Entrepreneurial Self: Fabricating a New Type of Subject*, S. Black (trans.), London: Sage.

Bröckling, U., Krasmann, S. and Lemke, T. (eds) (2010) *Governmentality: Current Issues and Future Challenges*, Abingdon: Routledge.

Cummings, M.L. (2015) 'Automation bias in intelligent time critical decision support systems', in D. Harris and W.-C. Li (eds) *Decision Making in Aviation*, Abingdon: Routledge, pp 289–294.

Dean, M. (1999). *Governmentality: Power and Rule in Modern Society*, London: Sage.

FCAS Forum (2021) 'Protocol FCAS forum'. Available from: https://www.fcas-forum.eu/protocols/protocol3/ [Accessed 4 January 2024].

Foucault, M. (2008). *The Birth of Biopolitics: Lectures at the Collège de France, 1978–79*. Basingstoke: Palgrave Macmillan.

Führungsakademie der Bundeswehr (2014) Auftragstaktik in der modernen militärischen Operationsführung – Ausprägung, Historie und Kritik: Ergebniszusammenfassung der Strategischen Analysen des Lehrgangs Generalstabs-/Admiralstabsdienst National 2012. Available from: https://doczz.net/doc/5709873/f%C3%BChren-mit-auftrag---f%C3%BChrungsakademie-der-bundeswehr [Accessed 6 January 2025].

French Ministry of Armed Forces (2019) 'Artificial intelligence in support of defence. Report of the AI task force'. Available from: https://www.defense.gouv.fr/sites/default/files/aid/Report%20of%20the%20AI%20Task%20Force%20September%202019.pdf [Accessed 4 January 2024].

Geyer, M. (2018) 'Der Unterführer als Feldherr im Taschenformat: Theorie und Praxis der Auftragstaktik im deutschen Heer 1869 bis 1945', *German History*, 36(1), 142–144. https://doi.org/10.1093/gerhis/ghx085

Government of the Netherlands (2023) 'Responsible Artificial Intelligence in the Military domain 2023 to Call to Action'. Available from: https://www.government.nl/documents/publications/2023/02/16/reaim-2023-call-to-action [Accessed 6 January 2025].

Hälterlein, J. (2021) 'Epistemologies of predictive policing: mathematical social science, social physics and machine learning', *Big Data & Society*, 8(1). https://doi.org/10.1177/20539517211003118

Hälterlein, J. (2023) 'Technological expectations and the making of Europe', *Science & Technology Studies*, 36(2): 26–46. https://doi.org/10.23987/sts.110036

Hälterlein, J. (2024) 'Imagining and governing artificial intelligence – the ordoliberal way: an analysis of the national strategy "AI made in Germany"', *AI & Society*. https://doi.org/10.1007/s00146-024-01940-0

High-Level Group on Artificial Intelligence (2019) 'Ethics guidelines for trustworthy AI', *European Commission*, 8 April. Available from: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai [Accessed 4 January 2024].

High-Level Group on Artificial Intelligence (2020) 'The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment', *European Commission*, 17 July. Available from: https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment [Accessed 4 January 2024].

Karakas, C. (2021) 'Defence industry cooperation in the European Union: rationale, initiatives, achievements, challenges', *European Parliament Research Service*. Available from: https://www.europarl.europa.eu/RegData/etudes/IDAN/2021/690607/EPRS_IDA(2021)690607_EN.pdf [Accessed 4 January 2024].

Keisinger, F. and Koch, W. (n.d.) 'Mission', *FCAS Forum*. Available from: . https://www.fcas-forum.eu/en/mission/ [Accessed 4 January 2024].

Klauke, S. (2021) 'Multi-Domain Combat Cloud – Infrastruktur und Innovationstreiber für europäische Wettbewerbsfähigkeit', in U. Lichtenthaler (ed) *Künstliche Intelligenz erfolgreich umsetzen*, Wiesbaden: Springer Fachmedien Wiesbaden, pp 15–39.

Koch, W. (2021a) 'On digital ethics for artificial intelligence and information fusion in the defense domain', *IEEE Aerospace and Electronic Systems Magazine*, 36(7): 94–111.

Koch, W. (2021b). 'On Digital Ethics for Artificial Intelligence in Hybrid Military Operation', NATO Science & Technology Organization Meeting Proceedings. Available from: https://www.sto.nato.int/publications/STO%20Meeting%20Proceedings/Forms/Meeting%20Proceedings%20Document%20Set/docsethomepage.aspx?ID=44861&FolderCTID=0x0120D5200078F9E87043356C409A0D30823AFA16F602008CF184CAB7588E468F5E9FA364E05BA5&List=7e2cc123-6186-4c30-8082-1ba072228ca7&RootFolder=%2Fpublications%2FSTO%20Meeting%20Proceedings%2FSTO%2DMP%2DIST%2D190 [Accessed 6 January 2025].

Koch, W. (2022) 'Elements of an ethical AI demonstrator for responsibly designing defence systems', in *2022 25th International Conference on Information Fusion (FUSION)*, Linköping, Sweden, 4–7 July, pp 1–8.

Koch, W. and Keisinger, F. (2020) 'Verteidigung und Verantwortung: Nutzung neuer Technologien in einem "Future Combat Air System"', *Behördenspiegel*, 36(4): 44.

Lemke, T. (2001). '"The birth of bio-politics": Michel Foucault's lecture at the Collège de France on neo-liberal governmentality', *Economy and Society*, 30(2): 190–207. https://doi.org/10.1080/03085140120042271

Lemke, T. (2021) *The Government of Things: Foucault and the New Materialisms*, New York: New York University Press.

NATO (North Atlantic Treaty Organization) (2021) 'An artificial intelligence strategy for NATO', 25 October. Available from: https://www.nato.int/docu/review/articles/2021/10/25/an-artificial-intelligence-strategy-for-nato/index.html [Accessed 4 January 2024].

Parasuraman, R. and Manzey, D.H. (2010) 'Complacency and bias in human use of automation: an attentional integration', *Human Factors*, 52(3): 381–410. https://doi.org/10.1177/0018720810376055

Rose, N., O'Malley, P. and Valverde, M. (2006) 'Governmentality', *Annual Review of Law and Social Science*, 2: 83–104. http://dx.doi.org/10.1146/annurev.lawsocsci.2.081805.105900

Saur, H. (2021) 'Multi-Domain Combat Cloud: a vision for the future battlefield', *Joint Air Competence Center*, May. Available from: https://www.japcc.org/essays/multi-domain-combat-cloud/ [Accessed 4 January 2024].

Skitka, L.J., Mosier, K.L. and Burdick, M. (1999) 'Does automation bias decision-making?', *International Journal of Human-Computer Studies*, 51(5): 991–1006.

US Department of Defense (2020) 'DOD adopts ethical principles for artificial intelligence', 24 February. Available from: https://www.defense.gov/News/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/ [Accessed 4 January 2024].

New Media, New Enemies: The Emergence of Automated Weapons in Counterterrorism

Jeremy Packer and Joshua Reeves

Introduction

This chapter argues that the creation of disciplinary-specific media technology automatically (and inevitably) produces military enemies, in that media produce new ways of perceiving the surrounding world and the threats that lurk therein. Our media-centric analysis rests upon several interwoven understandings of the recursive relationship between media, knowledge creation, and the production of enemies. At the most abstract level, we are suggesting that the existence of media necessitates and is necessitated by the production of enemies. The production of an enemy, following Deleuze's reading of Foucault's methodology, is a two-part act of making something visible or seeable and the capacity to name it as enemy; the power of the sayable (Deleuze, 1986). Following Friedrich Kittler, we approach media in terms of those technologies that specialize in the selection, storage, processing, and transmission of information (Kittler, 1999). From such a perspective, media encompass the techniques and technologies used to generate scientific knowledge such as the microscope, the military-produced technologies such as the global positioning system (GPS) used to guide missiles and Uber drivers, and the paper, ink and printing machines that have been spitting out daily newspapers for generations. Such an expansive definition has a long and prominent history (Innis, 2007; McLuhan, 1994; Peters, 2015) and is particularly relevant when considering the ways in which different disciplinary knowledges are dependent on the scientific instruments and modes of presentation used to generate their unique forms of knowledge (Daston and Galison, 2007). The production of international and domestic enemies, we emphasize, is as much a media-dependent process as the production of Mars Perseverance

Rover photographs. Without media like drone systems, radar and related intelligence, surveillance and reconnaissance technologies, these military enemies could not exist as such; these enemies are only intelligible to the extent that media technology introduces them to our scrutiny. Moreover, media are essential to the procedures used to destroy these enemies. As we see clearly illustrated on the battlefield in Ukraine, guided missiles and bombs, often shot by unmanned craft, rely on complex media systems in order to be 'guided' to their targets (Packer and Reeves, 2017). Automated weapon systems, utilizing cybernetic smart technology, have been key to both the Russian and Ukrainian efforts on land, air and sea.

Therefore, in the age of artificial intelligence (AI) machines, it is natural that automation has become essential to the identification of enemies on the battlefield. The US Department of Defense has a special category of information technology and intelligence which it calls I2 or *identity intelligence*: 'The intelligence resulting from the processing of identity attributes concerning individuals, groups, networks, or population of interest' (US Army, 2020: 40). Such intelligence and the turn towards increasingly automated forms of enemy detection are deeply rooted in military strategic and biopolitical formulations of threat and unity, war and politics. These roots, nurtured by scientific rationality, are deeply entwined in the epistemological/political desire to ferret out the hidden, latent and potential enemies of the state. Key to this process is the escalatory nature of positive feedback cybernetic systems: because they rely on continuous feedback loops, their epistemology is essentially dynamic. They are, as Antoine Bousquet puts it, 'chaotic': positive feedback cybernetic systems thrive on an 'amplification of disturbances' that constantly discovers new sites of potential intervention (Bousquet, 2009: 165). When applied in a political or military context, this means that enemies will always be found; with positive feedback systems, there is no way to ultimately find and neutralize all enemies; the system's operation demands the constant discovery of new problems to solve. This results in a recursive loop of detection and intervention which, ultimately, characterizes the way in which political/military enemies today become intelligible as enemies and are engaged appropriately. As we have argued previously: 'Every media revolution ushers forth new methods of slaughter' (Packer and Reeves, 2020: 9).

In this chapter, we will analyse how this media-dependent process has unfolded towards current trends in automated enemy detection; automated weapon systems, from this perspective, extend beyond the sphere of international military engagement to include AI weapons systems aimed at eliminating domestic political enemies – that is, to waging war on the internal enemy whose constant presence, as Foucault (2003) noted, is a necessary condition for the recognized existence of the external enemy. This is to say that producing enemies, both external and internal, has remained a constitutive element of the nation state.

We trace how three disciplines of knowledge production – biology, psychology and data science – have used media-centric methods of enemy production and enemy elimination (although, of course, we could have chosen other disciplines as well). During the past 150 years or so, the human sciences have repeatedly been turned to in order to produce unique media-specific methods for analysing and engaging military enemies and enemies of the state; today, of course, computer sciences have followed this trend. Often, these discipline-specific methods are mixed and synthesized, leading to hybrid disciplinary strategies for detecting biological (and often political) threats. For instance, biology has offered a number of methods for detecting enemyship within the human body – through a complex of technologies and procedures, the source of this biological enemyship was made visible and then subject to various regimes of biometric measurement and diagnosis. Given the specifically immutable and biological character of this enemy, it calls for very specific modes of engagement that, naturally, biology is equipped to provide. While this has given rise to methods for directly working on the species as a biological phenomenon (as in eugenics and less ominous health campaigns), it has also centred on physically isolating the biological threat – imprisonment, in particular, provides an excellent means of separating threatening bodies, in their presumed contagiousness, from the law-abiding population. The biological enemy, which is best detected by media like photographs, fingerprints, electrocardiograms (ECGs), brain scans, prenatal screening and other instruments of biological analysis, needs to be eliminated in a way suited to the biological – antiseptics, pharmaceuticals, exercise, surgery, confinement, detention, exclusion, castration, dismemberment, decapitation and genocide. In a word, biology is the problem; the solution is better biology.

Psychology, too, has its own instruments of enemy detection and its own methods of enemy elimination: the polygraph, for example, is a means of locating the deeply concealed psychological truth of the potential enemy. Psychoanalysis, the Rorschach test, battery tests and the other common methods of psychology approach the enemy as a specifically psychological phenomenon that is best detected by the traditional methods of psychological diagnosis. Yet once this enemy is detected, psychology does not specialize in biological forms of enemy engagement. While biological and psychological solutions to enemyship can be combined in various ways, particularly in penal and rehabilitative institutions, each has its own specific disciplinary commitments and methodological procedures – the physician might prescribe anti-depressants, while the psychologist places its confidence in therapy. Therefore, while various modes of *elimination* and *exclusion* are specifically biological solutions to enemyship, psychology offers its own media-specific projects – such as psychotherapy, mass propaganda and re-education – that address themselves to the imagined malleability of the human psyche. Enemyship expresses itself indelibly in the contours of the psyche; with good enough media technology, it can be detected and then addressed through diverse strategies of psychological correction. Psychology is the problem; the solution is better psychology.

Third, and perhaps most interesting, is the metadisciplinary field of data science and the focus on 'the network' as a ubiquitous, nebulous and all-encompassing field of contestation. In a trend that has truly taken off since the 2016 and 2020 US presidential elections, many people across the social sciences and humanities have come to regard 'the network' as the ultimate cause – as well as the ultimate solution – to the problem of the criminal, the domestic terrorist, the insurrectionist, the insider threat and the lone wolf. It is believed that 'The Internet' – especially as it is characterized by AI, pernicious social bots, and related agents of threatening organization and disinformation – is held to be not just the main way in which people are exposed to enemy attacks; it is, in a fascinating twist of irony, also held to be the ideal means of *combating* those attacks. From this point of view, social media platforms carry on an ambivalent political function: while they might allow extremists to recruit and propagandize the rest of us, William Roper, director of the US DoD's strategic capabilities office, recognizes that these platforms will also provide military and intelligence agencies with the best means to detect and fight back against its enemies: 'Data is going to be the fundamental fuel for national security in this century' (Strohm, 2016). Data science, it is claimed, will allow national security agencies and their corporate allies to combine biometrics, psychometrics and other disciplinary knowledge – especially from social science fields like psychology, criminology and communications – into a big data-driven portrait of who is an enemy, why they are an enemy, when their enemyship might express itself physically, and how to counter and prevent so-called radical conversions, disinformation and physical attacks in the future. The enemy, exists in, is altered by and responds to the network. The solution is a better, smarter network.

Above all, one trend is made clear by this media-specific analysis of enemies of the state: media make certain enemies newly visible, locate enemies that conform to their own unique standards of measurement, and then offer reprogramming resources that accentuate their own peculiar biases and capacities. This recursive loop of detection and intervention happens automatically, and in its most recent forms of cyber war or netwar, the process has become autonomous. Such processes correspond to different ways of knowing and valuing the world, and it therefore illustrates the politically fraught ways in which media systems identify, classify and mark individuals and collectivities (Robertson, 2012; Silva, 2016). These different media-specific modes of enemy engagement have different commitments to solving 'the enemy within' – that is, the domestic enemy within the nation, in particular the latent enemy that exists deep within us all. As these competing modes of detection, analysis and correction come to realize that we are all a potential threat – that the latent violence in us all is just waiting to be awakened by vodka, bullying or the right TikTok video – the question revolves around how to best locate and suppress these potentialities through autonomous means. Of course, this applies above all to AI and related forms of contemporary enemy identification: as the 'chaotic' nature of automated enemy production becomes clearer, the network's methods represent a grim escalation in this process. Given the now more or less compulsory nature of digital participation in social and economic life – particularly when that compulsory participation is combined with systems of ubiquitous data capture (Borradaile and Reeves, 2020) – we all, as potential enemies, comprise both a dynamic database and a field of experimental military intervention.

1. Enemies in the blood: biometric identification systems

Biology was one of the earliest disciplinary sites for analysing threats internal to a population. While the police force gained power and resources throughout the 19th century, its need for new knowledges and new instruments grew. As eugenics crept into many nooks of scientific inquiry, many scientific experts and community leaders argued that police authorities could address crime rates by linking citizens to vast surveillance databases filled with photographs, family genealogies and life histories. These databases were in part aimed at providing authorities with a vast resource they could use to isolate the genesis, habits and physical characteristics of *Homo criminalis*, the biologically determined criminal being (Beirne, 1993).

Best known for his work in establishing eugenics, Francis Galton was also a photography expert and the inventor of composite photography. For Galton, composite photography could help highlight the essential characteristics of a given population group: 'Having obtained drawings of photography of several persons alike in most respects, but differing in minor details, what sure method is there of extracting the typical characteristics from them?' (1878: 97). This process of 'extracting the typical' was used to theorize the average physical characteristics of ideal and threatening populations. For one of his first experiments, Galton made a composite photograph of several murderers and violent thieves. According to him, this composite process smooths away the unique facial features of the individual offenders and instead discloses something essential about 'the criminal'. In the composites: 'The special villainous irregularities [of the individuals] have disappeared, and the common humanity that underlies them has prevailed. They represent, not the criminal, but the man who is liable to fall into crime' (1878: 97–98).

Galton's composite photography was designed to shift the plane of enemy intelligibility: while each individual photograph reveals the superficial visibility of a criminal, the composite locates a threat potentiality that lies beyond the level of immediate visibility – something that is only detectable with the collection, storage, comparison and synthesis of other archived photographs (see Sekula, 1992). The size and shape of one's facial features; one's weight; one's height and one's biological make-up, as analysable through the photograph and related technologies of measurement and capture, revealed an abundance of information about one's potential to threaten the social order (cf. Siegel, 2014: 206–207).

This early manifestation of eugenic science gave birth to a whole host of biometric technologies designed to discover and analyse the specifically biological conditions of enmity: the race science of the early 20th century, in particular, thrived on measuring and classifying individuals according to racial and class types (and then, of course, assessing their threat to the social body accordingly; see Pick, 1993: 75–87). Therefore, the biological understanding of enmity forwarded by eugenics was founded on a media-specific logic of enemy identification. Its primary method of representation, photography, was accompanied by an entire apparatus of measurement, storage and analysis – including callipers, ink pens, rulers, gauges and filing cabinets (Reeves and Packer, 2013; Morris-Reich, 2016: 34–84). Located within this media apparatus, the photograph takes on a specific representational function; its significance is always deferred, always figural, in that it succeeds only in continuously circulating the physiological essence of the criminal. The individual represented in the photograph disappears; the photograph, just like all the anonymous files surrounding it, discloses nothing but *Homo criminalis*. It carries with it every scoundrel in the history of humankind. As such, the enemy in the photograph is a speculative enemy – the enemy as sign, as spectrum, as endless circulation of the natural born killer.

However, with computerization, the main function of these biometric technologies shifts from the speculative to the investigative. At that point, the photograph becomes more than just a sign filed away in a stack of other redundant signs. While fingerprints have been gathered and stored by police departments since the early 20th century, pre-digital fingerprint analysis was extremely difficult and painstaking, not to mention worthless in most investigations. But with the formation of computerized databases in the 1980s and 1990s, biological data came to comprise the very backbone of criminal identification (Magnet, 2011: 51–68). Computerized biometrics became a media-specific form of threat analysis that allowed for the transformation of bodily traces and body representations into easily processable data. Unlike the composite photograph, the systems of classification specific to computerized biometrics did not search for the general in the particular; instead, they sought to capture the molecular biological peculiarities of the individual. This individual no longer signified 'the murderer'; the murderer, who had lost its photographic essence, could only be discovered by the vast accumulation of biological data about the guilty (as well as the innocent). The human biomachine, which oozes a neverending datastream from its orifices and pores, was constantly compelled to testify as to its own guilt.

Intensive media (those used to make legible the inside of a body) open subjects up to friend-enemy assessment and draw into the realm of analysable data an ever-growing temporal frame (see Murray, 2009: 69–87). Dangerous data are 'flagged,' cross-referenced and acted upon. Civilizational affiliation, family lineage, religious genealogy and cultural heritage are brutish markers of enemyship. Slightly more refined media are used to build lifelong datasets for tracking changes in attitudes, beliefs, political affiliations, infirmities, access to weapons and proximity to ideological contagion. Fine-tuned media measure changes in bodily function (brain waves, perspiration, breathing patterns, muscular ticks, eye movement and body temperature) not immediately accessible to human perception and are indicative of more immediate threats. These fine-tuned threat detection practices, of course, have served an important function in foreign US war zones like Afghanistan, where local populations are subjected to various forms of biological analysis – like fingerprint impressions, iris scans, tissue sample analysis (Gorman, 2011) and even gait evaluation (Singh et al, 2021) – and then compared with biometrics-enabled watchlists (Center for Army Lessons Learned, 2011). For the US Army, mapping 'the human terrain' (González, 2008) is an essential step in fighting the internal enemy by analysing each encountered person and classifying them, according to threat databases, in terms of their threat potential. These same methods of analysis, especially under the guise of facial recognition, are being deployed against the domestic insider threat (Gates, 2011; Andrejevic and Selwyn, 2022); indeed, at least half of all US adults are stored in the FBI's facial recognition database (McLaughlin, 2016). This vast human terrain – which, as we will see in a moment, takes on an especially interesting character vis-à-vis the network – derives its investigatory potential from the presumed permanence and stability of the features captured. In a word, digital biometric analyses do not rely on waist sizes, but on those less malleable elements of the human terrain such as the iris signature and facial structure. This enemy cannot be captured in a composite; it is an enemy deprived of its community, its family tree of incurable delinquents. Moreover, the biological enemy produced by computerized biometrics is a *whole* enemy. Its traces do not gesture towards other criminals, insurgents or terrorists, but only towards its own indivisible identity. It circulates in an unmistakable digital monism. Even if the terrorist cuts off its own thumb or gouges out its own eye, it still remains a terrorist. The biologically determined enemy must be detected and removed from circulation. It must be taken out of the data stream.

2. Psychometric identification systems

The psychologically determined threat, on the other hand, introduces different considerations. Psychological analysis has always in one way or another been driven by the methodological conditions made allowable by media technology. As Friedrich Kittler explained, the phonograph was central to Freud's development of psychoanalysis and his 'talking cure' (1999: 141). Peter Galison (Galison and Packer, 2016) provides a media-centric explanation for the popularity of the Rorschach test, arguing that it was thought to function like an 'X-ray of the soul' (Johnson 2012), thus providing a clear mediation of the previously invisible unconscious. Hermann Rorschach invented the test and considered it a technology of 'psychodiagnostics' (1942). This was given such epistemological credence that it was eventually used to test Nazi war criminals at the Nuremberg Trials.

The US military did not extensively turn towards psychology until the First World War, when it used psychological profiling as part of its newly implemented Selective Service process. In addition to assessing the psychological wellbeing of potential soldiers, the military carried out a particularly notable set of intelligence tests, Army Alpha and Army Beta, on more than a million of their own draftees to locate mental capacity and deficiency. In point of fact, it could be argued that American involvement in the First World War also had a profound effect on legitimating the floundering young discipline, as it produced the first significantly large psychological database from which broad societal generalizations began to be made (Samelson, 1977). Further, the use of the Hollerith Tabulating machine to compute large quantities of data and store this data on punch cards had a broad set of ramifications for how many different forms of variable analysis could be run. Mental 'deficiency' within the draftee population was generalized to the broader population. Army Alpha and Army Beta were spearheaded by Harvard psychologist and eugenicist Robert Yerkes, who used his research on soldiers to substantiate such claims as 'no one of us as a citizen can afford to ignore the menace of race deterioration' (see Tucker, 1996: 82).[1] We can think of this in terms laid out by Foucault (2003) in which the 'race' war works to cleanse the internal population of the 'unfit', while also legitimating attempts to characterize the national claims for warfare by understanding the other (the fascist, the communist, the terrorist and so on) as psychologically deficient, unfit and 'dangerous'. With this merging of military and domestic purposes, psychology found increasing legitimacy for its capacity to assess individuals and populations during this period, in large part because of its supposed success in the war effort. The workings of the

[1] For a fuller account of this long and pained history, see Richards (2003).

unconscious came to be seen as something only very partially understood, but also something which, if properly excavated, would provide immense political and military value.

Understanding the psychology of fascism and communism became common for the field of psychology during the 1930s. Wilhelm Reich's *The Mass Psychology of Fascism* (1933/1980) was the most famous and it described the rise of fascism and Bolshevism as both resulting from the same psychological impulses and a latent openness to authoritarianism. The specificity of the analysis is far less interesting (though Deleuze and Guattari reanimate it in *Anti-Oedipus*, 1972/1983) than the general principle upon which Reich's assessment rests. Essentially, for Reich, political will is always open to psychological misdirection. Psychological manipulation can directly lead to individuals and, more importantly, 'the masses' being led to act against their own interests (à la Marxist critique) by potent symbols that are transmitted via mediation. For Reich, the swastika is one such talisman which was used to manipulate the unconscious by breaking directly into the primal scene. The human psychological condition, for Reich, is something of a tabula rasa that can be easily overwritten. The unconscious is imagined as a screen on which overpowering ideologies can be projected.

In the postwar period in the US, psychology was given the task of helping to root out overt communists and latent communist tendencies. Andrew Stouffer, whose Second World War research resulted in the four-volume text *Studies in Social Psychology in World War II: The American Soldier*, took up a related examination during the Cold War: analysing the American character for signs of susceptibility to communism and extremism. His 1955 book *Communism, Conformity, and Civil Liberties: A Cross Section of the Nation Speaks Its Mind* argued that Americans were not overly concerned regarding communism. Knowledge of the soldier remained linked to knowledge of the enemy. More broadly, Stouffer's book addressed the question for the day, which was to what degree Americans are susceptible to radical and intolerant thought. This concern continued and psychology has been used to regard other forms of internal threat to the American psyche such as a worry over the 'paranoid style' of thinking (Bratich, 2008) and, more recently, over terrorism. This sort of social psychology did not depend on more elaborate media technology than that of its precursors. Tape recorders and file cabinets, which stored interviews and rudimentary statistical data, dominated the field of social psychology.

At the level of the individual communist sympathiser or the outright Soviet spy, psychology helped to provide new media for making the subject 'speak the truth' of its treason. Polygraph machines became prominent features of Red Scare and the House Un-American Activities Committee (HUAC) interrogations. More elaborate schemes for applying the field's knowledge to the problem of truth telling involved LSD, hypnosis and various modes of psychological torture. During this era, a host of media were applied in the realm of counterespionage to test citizens' adherence to US doctrine. The lie detector, for example, was used during the Red Scare to examine subjects for Communist/Soviet ties (Segrave 2004: 48–72). In addition, a great deal of work was done by the Central Intelligence Agency (CIA) and its global competitors to elaborate new forms of psychological analysis. For instance, during the early 1950s, the field of psychology was invested in studying and elaborating new means for carrying out *menticide*, which, according to the *Random House Dictionary*, is 'the systematic effort to undermine and destroy a person's values and beliefs, as by the use of prolonged interrogation, drugs, torture, etc., and to induce radically different ideas'. Menticide was seen as a legitimate end goal of psychological practice and as a military strategy of utmost importance (Robin, 2001: 163–172). A 1956 CIA document describes in great length what psychological techniques were being successfully used in the Soviet Union and China which might be applied equally well by US security forces (CIA, 1956). Further, knowledge of these techniques could be reverse engineered in order to prepare defences against them. The final section of the 122-page report focuses on a 'Theoretical Analysis' of the various tactics. Tellingly, the report describes each interrogation or torture as an 'experimental situation' (1956: 114). Each interrogation is a psychological experiment in the possibilities of the field to weaponize itself. With no oversight boards to worry about and driven by the brute necessity of the Cold War, these interrogations push the limits of experimental practice. They supposedly open up the enemy to such a level of sophisticated reprogramming that they cease to be an enemy at all. Further, psychology will be used extensively during the Cold War to ferret out Soviet spies and communist sympathizers. At that juncture, the power of psychology was further elaborated through a host of chemical, technical and bodily transformations to unveil hidden knowledge and to actively 'flip' the spy –to reprogramme them to create the double agent. Psychology was seen as the scientific means for undoing or decommitting enemies.

More recently, with the rise of the War on Terror, a bevy of psychologists have come to the aid of the war effort to delve into the psychological make-up of the terrorist. We see again and again the renewal of faith in a discipline devoted to understanding the workings of the mind in a manner that allows for an assessment of the cause: 'What makes a terrorist?' Yet one of the underlying considerations into psychology as a method for understanding the adversary is an unwritten assumption that the adversary, by definition, suffers from some condition that makes them different, aberrant, dangerous, unhinged, unmoored, unstable and fundamentally irrational (see Foucault, 2013). The confounding problem that tends not to be addressed is how it is that the 'friend' is never placed under the same form of scrutiny. Using the tools of psychology to discover enemies already assumes that enemy-ness is a discoverable trait as opposed to a constructed category created to legitimate warfare and inequity.

Over the past two decades, psychology has had to redirect its focus in order to be seen as a useful anti-terrorist science. A new war demands a new psychology. And new psychology demands new media (see Schuurman and Elijkman, 2015). As a number of well-cited terror experts in the field of psychology will attest, we have duelling, or perhaps innumerable, latencies which can supposedly be exploited by powerful media-enabled political agents. The work of John Horgan, in particular, stands out in that over the course of three separate books, he analyses the psychology of terrorism (2014), 'walking away' from terrorism (2009), and 'leaving terrorism behind' (Bjorgo and Horgan, 2009). In short, the role of psychology is not only to locate the psychological mechanisms that lead to terrorism, but psychology and its signature media should also play a central role in the battle against the enemy.

3. Enemies in the network

> From phony news on websites to terrorist propaganda on social media to recruitment videos posted by extremists, conflict in the information domain is becoming a ubiquitous addition to traditional battlespaces. Given the pace of growth in social media and other networked communications, this bustling domain of words and images – once relegated to the sidelines of strategic planning – is poised to become ever more critical to national security and military success around the globe. (Wade Shen, DARPA, Quantitative Crisis Response [QCR])

Our analysis of biology and psychology points out the degree to which a given field of study overestimates its capacity to locate the true nature of enemies and the root cause and site of their abnormality. Further, the solutions they provide are overdetermined by their disciplinary focus and their modes and media of enquiry; of data collection, storage and processing. Biology was given to analysis of the body and blood, and psychology to the mind, psyche or brain. In this final section we will examine the way in which data science (and its subdiscipline network science) approaches 'the network' as the new locus for understanding enemies and creating solutions. While one can say that a body or brain exist as such, they are still produced by biology and psychology as a particular kind of object to be understood according to the rules of their field. Brains are both materially and discursively produced, and so too is the network. While there has been extensive academic work done on the meaning, history and importance of networks, network societies and network power (Castells, 1996; Mattelart, 2000; Galloway and Thacker, 2007), it is the collective works produced for the RAND Corporation by Arquilla and Ronfeldt that have set the stage for analysis and theoretical development when it comes to threats against state power and the state formation from and by networks. These RAND documents include, but are not limited to, *Networks and Netwars: The Future of Terror, Crime, and Militancy* (2001), *Swarming and the Future of Conflict* (2000), *The Emergence of Noopolitik: Toward an American Information Strategy* (1999), *Strategic Appraisal: The Changing Role of Information in Warfare* (1999), *The Zapatista 'Social Netwar' in Mexico (1998), In Athena's Camp: Preparing for Conflict in the Information Age* (1997), *The Advent of Netwar* (1996) *and Cyberwar is Coming!* (1993). This outpouring of analysis continues to orient much of the discourse around the centrality of networks as the mechanism by which post-Cold War threats arise, hide, grow, are known and are excised. The network remains a concept ripe for data scientific production and political machination.

Importantly, the network it is claimed is more than just a battlespace – it is imagined to be the cause and the solution to the current crisis in 'identity intelligence' and it is said to exhibit unique criminogenic qualities, most importantly 'Complex Global Microstructures' (Taylor, 2015: 98). Arquilla and Ronfeldt (1996) – along with Galloway and Thacker (2007) – suggest that network power and network politics operate in asymmetrical ways. Yet they also operate in a recursive fashion – that is, the network produces the enemy of the network, which then produces a network solution to the enemy problem. Thus, we are interested in how the network functions in a similar fashion to the psyche in psychological discourse or blood in an older biological discourse. This is to suggest that it functions as a 'free-floating legitimator' (Packer, 2010) used to substantiate all sorts of truth claims. The supposed power of the network to radically alter how humans act and to what political, ideological and religious causes to which they adhere is particularly prominent. As such, the network is widely described as a site of struggle where netpower runs unrestrained.

The network is said to see the with instability and unrest, but ultimately, according to the US's Defense Advanced Research Projects Agency (DARPA) and Google alike, the network can be made to automatically and autonomously take care of itself. A host of autonomous programs publicly known – and some clearly unknown – continue to be unleashed into the network ecosphere in an attempt to locate what we are calling *homo inimicus* – the malleable humans who reside in the network as an always potential threat. While these projects bear some telling similarities with previous attempts to use media to unveil threats, in other ways the internal logic of this particular media system considerably alters the methodological means and epistemological convictions of previous modes of enemy detection. Our concern here is not necessarily whether network power does indeed produce terrorists or, more to the point, whether the network actually has power. Instead, we posit the notion of netpower as a claim that networks are overwhelmingly powerful – so much so that they are said to be able to reprogramme a subject's political convictions so thoroughly that they feel obliged to commit acts of mass violence. As DARPA's Quantitative Crisis Response (QCR) team concisely explains: 'automated digital tools … can help operational partners better understand how information is being used by adversaries and to quantitatively predict and assess – in real time and at scale – the effects of those campaigns and of countermeasures' (Shen, 2017). Therefore, the automation power of the computer network is assumed to have the power to modify the cognition and behaviour of its targets.

One vexing problem for all forms of knowledge production, including network-focused data science, is the set of one. The singular, seemingly unattached and hence difficult-to-locate lone wolf serves as a very telling example of how the network functions. When a lone wolf is said to act on its own for often unclear purposes, the question for *identity intelligence* becomes one of looking for the presence of affective engagement with 'terrorist propaganda' or other mechanisms of 'radicalization' when it is not clear that any may exist. These lone wolves are seen as 'DIY terrorists' who are triggered by the network into action. One example that has supposedly amplified the message is a May 2016 statement by ISIS leaders that everyone living in Europe and the US is an enemy: 'There is no such thing as innocents' (Tucker, 2016). Impetus and tactics both circulate in the network (Derick et al, 2016). The DIY terrorist is not trained by specialists or given tactical advice by superiors, but is nonetheless empowered by the network for successful terrorist action.

It is unclear whether netpower activates something inherent in the lone wolf or creates them entirely. In either case, the lone wolf as a category is problematic because of its assumed singularity and uniqueness. To detect the lone wolf through analysis suggests commonality. Is it alone or part of a different pack arrangement whose togetherness exists through the network, however tenuously? How many strands of connection, from how far away, and for how long must they exist to activate togetherness or a shared goal? This is to ask the following: is the lone wolf a predictable outcome of netpower? And if it is, what tools could be used to hunt down the lone wolf before they are activated?

To a great degree, this is the kind of problematic for which DARPA is attempting to engineer a solution via QCR. Following a lone wolf attack in Norway in 2011 that led to 77 deaths, law enforcement officials assumed that it was impossible to use traditional digital technologies to detect and deter a lone wolf attack before it happened. Yet when faced with the same question, four members of the Swedish Defense Research Agency suggested that the difficult to locate 'weak signals' might be operative that if detected could provide predictive clues as to the potential activation of lone wolves (see Kaplan, Lööw and Malkki, 2014). Other research (Meyer, 2013) has attempted to tackle the same situation and has come to much the same conclusion. Even if these 'weak signals' existed and were present and detectable in theory, a temporal barrier in computational speed and an access barrier to data selection would hinder fully predictive capacities in real time. The data needed are often located in commercial proprietary systems that are not fully available to security forces, and the amount of data is too vast to run the necessary algorithms on everyone at all times.

Even so, we might ask the following: are 'weak signals' any easier to locate than lone wolves? In most cases, the work being done is forensic recreation in which a pattern of activities that might be indicative and suggestively causal are located after the fact. Traces of activity – a lone wolf's search history and Facebook posts, for instance – are used to piece together a theoretical chain of events that might have led to the lone wolf's activation. This *ex post facto* forensics endeavour is itself used to legitimate netpower's own existence and strength: these traces are regarded as 'weak signals' that provide potential insight into future wolves. This leads to the compilation of signals that might guide the search for becoming-wolves.

Describing Jigsaw, Google's anti-terror initiative, a *Wired* journalist remarked:

> Jigsaw, the Google-owned tech incubator and think tank … has been working over the past year to develop a new program it hopes can use a combination of Google's search advertising algorithms and YouTube's video platform to target aspiring ISIS recruits and ultimately dissuade them from joining the group's cult of apocalyptic violence. (Greenberg, 2016)

Jigsaw not only tries to locate potential lone wolves, but it then attempts to use netpower to undo the work of netpower. Jigsaw reterritorializes the purportedly natural pathways that a potential lone wolf travels in the network by altering how its search engine works and guiding becoming-lone wolves towards anti-Jihadist websites, videos and communities. Likewise, with its Moonshot CVE program, Google began in 2016 applying 'start-up thinking to the field of countering violence extremism' (Hooper, 2017). It has continued to fund and develop dozens of network solutions to problems which they associate with extremism. For instance, one program called Perspective 'is an API that uses machine learning to spot abuse and harassment online' (Civic Tech Field Guide, 2023). In Perspective, conversation AI is 'studying how computers can learn to understand the nuances and context of abusive language at scale' (Jigsaw, 2017). Each month Moonshot provides its Threat Bulletin, which outlines that month's most pressing concerns. In August of 2023 this included, 'Anti-Government and Anti-Authority Violent Extremism (AGAAVE), Conspiracy theories, Domestic violent extremism, Mass Shooting … Racially or Ethnically Motivated Violent Extremism, Serious violent crime, Violent dissident Republicanism' (Moonshot, 2023). The Threat Bulletin is meant to inform partners and the public of which threats had high incidence in the network during the previous month. In the case of conspiracy theories, to choose one from that month's Bulletin, issue five of its online journal *The Current* was devoted to conspiracy theories. Each issue of *The Current* is divided into three sections which outline the problem, the perspective and what countermeasures have been developed to operate in the network. Four countermeasures are presented; fact checking, managing harmful content, the redirect method, and prebunking. Fact checking embeds a Google misinformation warning into searches. Managing harmful content is a standard content moderation program used by Google to flag and erase harmful content from its sites such as YouTube. Prebunking provided top-of-page information that would legitimate institutions, technologies or historical claims that were the topic of a conspiracy further down a page. The redirect method would simply redirect search results and page content away from sites and links deemed dangerous towards counternarratives. In each of these cases, AI is used to locate, track and in some cases alter netpractice in order to harness netpower.

A second and related netpower enemy is the insider threat, the rogue. Edward Snowden stands as the exemplary figure, but others such as Chelsea Manning, who leaked some 500,000 confidential military documents in 2010, are also seen as rogue agents. New 'insider threat technology' is being developed to locate and counteract the rogue (Sternstein 2016a). The rogue is one who started as a friend, but turns against the military and uses their previous insider status as a lever for unveiling a weakness in the military's structure or public relations. The movements of the rogue can be reverse engineered, largely because their movements occurred within the subnetwork of the military's communications infrastructure. This reverse engineering acts to reconstruct the pathways and activities that both made them go rogue and gave them the means for doing so. In other words, the rogue is produced in the specialized military network and as such is even more of a threat to that very system. They are inhabitants of the system that is supposed to be secure from outside penetration, but as they become rogue, they expose the openness of the system to external forces or self-destructive internal forces. Their existence and the reverse engineering are used to remake them as an analysable historical entity. Actuarial predictive models can then be generated to assess other members of the network with a 'FICO[2]-like score for integrity' (Sternstein, 2016b) in order to constantly

[2] Fair Isaac Corporation (FICO) created this credit score so that lenders can use borrowers' FICO scores to assess credit risk.

monitor and index all sub-networkers' potential to go rogue. As a parting gift to the National Security Agency, Snowden left them his digital trace – which today is used to create I2 models for shoring up military network security and identifying future insider threats. In the cybersecurity industry we see the development of numerous multi-pronged and techno-automated modes of 'insider threat technologies'. While not specific to biopolitical goals, these forms of automated threat detection and elimination take the network itself as the 'friend' which needs protecting. The businesses and institutions that contract such services are the inherent friend, but ensuring the ongoing viability of network power is the ultimate goal. Here is a short list of what one suite of deception detection products focus on:

- Deception decoy servers and devices, file shares, credentials, files, databases and other decoy elements which are intermingled within the production environment and alert upon engagement.
- Application server decoys for detection of reconnaissance or alerting on the use of authorized credentials in unauthorized ways.
- Active Directory protection, which identifies unauthorized queries, diverts attackers to decoys, and delivers the attacker misinformation to derail their attack
- Deception for native cloud technologies, such as containers, serverless functions and storage buckets
- Through the use of its machine learning, visibility to exposed credentials that create attack paths and access as well as unauthorized devices added to the network.
- Identification and substantiation of policy violations or malicious activity (*Business Wire*, 2019).

While the desire to locate the insider threat via an engorged corporate security industry continues to thrive, several new modes of automating the search for enemies have emerged that draw from the biometric, the psychological and network analysis. More broadly, the tech industry responded to criticism from governments across Europe and North America to form the Global Internet Forum to Counter Terrorism (GIFCT), which according to its own PR, 'brings together the technology industry, government, civil society, and academia to foster collaboration and information-sharing to counter terrorist and violent extremist activity online' (GIFCT, 2023). It works with a number of research, policy and technology partners, including the Global Network on Extremism and Technology. One primer on AI's relevance to combating extremism provides an overview of the myriad ways in which AI is posited as a tool in combating terrorism. The stated goal for the use of AI is the creation of 'a 'healthy' online space, free from terrorist content, propaganda material and fake engagement (Schroeter, 2020). The notion of an information 'jungle' highlights the neocolonialist rhetoric we might expect from an organization working to combat 'insurrectionists' and 'terrorists'. Stuart Hall (1981) explained how the notion of the jungle functions as a mechanism for naturalizing white supremacist ideology. The notion that technology and civilization are under threat from the untamed jungle animated the CIA psychological discourse from the 1950s and continues to flourish in the 2018 book *The Jungle Grows Back: America and Our Imperilled World* (Kagan, 2018).

The tactical, strategic and absurdly profitable control over 'search' is in key ways the foundation upon which the internet functions (Zuboff, 2019). Search engine algorithms are not unlike GPS, which was initially used to guide Inter-Continental Ballistic Missiles (ICBMs); we might better think of them as 'search and destroy' engines. The remodulation of such guidance would redirect users away from 'extremist' or 'radical' content, and instead towards 'moderate' and 'balanced' content. While the development of tactical military media have been used to locate, surveil and guide munitions towards enemies, internet or network searches are aimed not so much at enemies, but at content and engagement patterns that might infect the network. Again, the recursive logic teaches us that not only do individuals function as patterns and nodes in the network; they are also *outputs* of the network.

Conclusion

The network has shown that the enemy is not the enemy. The enemy is a trace. Biometric identification systems, psychometric identification systems, algorithmic network analysis – these current forms of insider threat detection direct themselves to the molecular. Terrorism might be innate, but only as a capacity, an inclination, a sleeper threat that can be awakened at any moment given the right economic, social or cultural circumstances. Ultimately, the predominant security apparatuses in the US and Europe recognize that, somewhere deep inside, we all have the capacity to be 'with the terrorists'. These security apparatuses and their corporate allies are therefore directing strategies towards finding what awakens that explosive potential. Unlike dystopian libertarian nightmares to the contrary, the future methods of neutralizing the domestic security threat are unlikely to involve concentration camps and machine gun squads. This form of old biopolitics is not well suited to domestic liberal population management; and it is not really necessary, at least in the near future. Instead, a more subtle and fine-tuned political regulation of populations will more likely involve the security apparatus – in league with its corporate partners like Google – detecting our threat potentiality, altering our media exposure, recalibrating our political inclinations and culturally suppressing our innate capacity for becoming enemy. This clarifies the new affordances of AI and 'the network' in their capacities as automated weapons in domestic wars on dissent and rebellion. With emerging technologies of detection and cultural governance, we are the threat, the database and the experiment.

References

Andrejevic, M. and Selwyn, N. (2022) *Facial Recognition*, Cambridge: Polity Press.

Arquilla, J. and Ronfeldt, D. (1996) *The Advent of Netwar*, Santa Monica, CA: RAND.

Beirne, P. (1993) *Inventing Criminology: Essays on the Rise of 'Homo Criminalis'*, Albany: State University of New York Press.

Bjorgo, T. and Horgan, J. (2009) *Leaving Terrorism Behind: Individual and Collective Disengagement*, New York: Routledge.

Borradaile, G. and Reeves, J. (2020) 'Sousveillance capitalism', *Surveillance & Society*, 18(2): 272–275. http://dx.doi.org/10.24908/ss.v18i2.13920

Bousquet, A. (2009) *The Scientific Way of Warfare: Order and Chaos on the Battlefields of Modernity*, New York: Columbia University Press.

Bratich, J. (2008) *Conspiracy Panics: Political Rationality and Popular Culture*, Albany: State University of New York Press.

Business Wire (2019) 'Cyber deception ranked top choice for insider threat detection; Attivo Networks® awarded for solution efficacy', 20 November. Available from: https://www.businesswire.com/news/home/20191120005189/en/Cyber-Deception-Ranked-Top-Choice-for-Insider-Threat-Detection-Attivo-Networks%C2%AE-Awarded-for-Solution-Efficacy [Accessed 12 January 2024].

Castells, M. (1996) *The Rise of the Network Society*, Oxford: Blackwell.

CIA (Central Intelligence Agency) (1956) *Communist Control Techniques: An Analysis of the Methods Used by Communist State Police in the Arrest, Interrogation, and Indoctrination of Persons Regarded as 'Enemies of the State*, 2 April, declassified September 2000. Available from: https://www.cia.gov/readingroom/docs/CIA-RDP78-03362A000800170001-2.pdf [Accessed 12 January 2024].

Center for Army Lessons Learned (2011) *Commander's Guide to Biometrics in Afghanistan: Observations, Insights, and Lessons*, Fort Leavenworth, KS: CALL.

Civic Tech Field Guide (2023) 'Perspective', *Civic Tech Field Guide*, 20 February. Available from: https://directory.civictech.guide/listing/perspective [Accessed 12 January 2024].

Columbia University Computer History (n.d.) 'Hollerith 1890 census tabulator and the evolution of the IBM punch card'. Available from: http://www.columbia.edu/cu/computinghistory/census-1890.jpg [Accessed 20 February 2024].

Daston, L. and Galison, P. (2007) *Objectivity*, New York: Zone Books.

Deleuze, G. (1986) *Foucault*, Frankfurt am Main: Suhrkamp.

Deleuze, G. and Guattari, F. (1972/1983) *Anti-Oedipus: Capitalism and Schizophrenia*, Minneapolis: University of Minnesota Press.

Derrick, D.C. et al (2016) 'Ideological rationality and violence: an exploratory study of ISIL's cyber profile', *Dynamics of Asymmetric Conflict*, 9(1–3): 57–81. http://dx.doi.org/10.1080/17467586.2016.1267866

Foucault, M. (2003) *Society Must Be Defended: Lectures at the College de France, 1975–76*. G. Burchell (trans.), New York: Picador.

Foucault, M. (2013) *Madness and Civilization: A History of Insanity in the Age of Reason*. R. Howard (trans.), New York: Vintage.

Galison, P. and Jeremy P. (2016) 'Abstract materialism: Peter Galison discusses Foucault, Kittler, and the history of science and technology', *International Journal of Communication*, 10: 3160–3173.

Galloway, A.R. and Thacker, E. (2007) *The Exploit: A Theory of Networks*, Minneapolis: University of Minnesota Press.

Galton, F. (1878) 'Composite Portraits', *Nature*: 97–100.

Gates, K.A. (2011) *Our Biometric Future: Facial Recognition Technology and the Culture of Surveillance.*, New York: New York University Press.

Global Internet Forum to Counter Terrorism (2023) 'Home Page', *Global Internet Forum to Counter Terrorism*. Available from: https://gifct.org/ [Accessed 12 January 2024].

González, R.J. (2008) '"Human terrain": past, present, and future applications', *Anthropology Today*, 24(1): 21–25.

Gorman, C. (2011) 'How biometrics helped to identify the master terrorist', *Scientific American*, 2 May. Available from: https://www.scientificamerican.com/article/how-biometrics-helped-to-identify-master-terrorist/ [Accessed 12 January 2024].

Greenberg, A. (2016) 'Google's clever plan to stop aspiring ISIS recruits', *Wired*, 7 September. Available from: https://www.wired.com/2016/09/googles-clever-plan-stop-aspiring-isis-recruits/ [Accessed 12 January 2024].

Hall, S. (1981) 'The whites of their eyes: racist ideologies and the media', in G. Bridges and R. Brunt (eds) *Silver Linings: Some Strange Strategies from the Eighties*, London: Lawrence & Wishart, pp 28–52.

Hooper, S. (2017) 'Exclusive: Google working with UK on counter-extremism', *Middle East Eye*, 28 March. Available from: https://www.middleeasteye.net/news/exclusive-google-working-uk-counter-extremism [Accessed 12 January 2024].

Horgan, J. (2014) *The Psychology of Terrorism*, New York: Routledge.

Horgan, J. (2009) *Walking away from Terrorism*, New York: Routledge.

Innis, H. A. (2007) *Empire and Communications*, Lanham, MD: Rowman and Littlefield.

Jigsaw (2017) 'Conversation AI', *Jigsaw*. Available from: https://jigsaw.google.com/projects/#conversation-ai [Accessed 12 January 2024].

Johnson, C.Y. (2012) 'What the Rorschach tells us', *Boston Globe*, 10 June. Available from: https://chsi.harvard.edu/files/chsi/files/what_the_rorschach_tells_us_copy.pdf [Accessed 12 January 2024].

Kagan, R. (2018) *The Jungle Grows Back: America and Our Imperiled World*, New York: Alfred A. Knopf.

Kaplan, J., Lööw, H. and Malkki, L. (2014) 'Lone wolf and autonomous cell terrorism', *Terrorism and Political Violence*, 26(1): 1–256. https://doi.org/10.4324/9781315724263

Kittler, F. (1999) *Gramophone, Film, Typewriter*, Palo Alto: Stanford University Press.

Magnet, S. (2011) *When Biometrics Fail: Gender, Race, and the Technology of Identity*, Durham, NC: Duke University Press.

Mattelart, A. (2000) *Networking the World, 1794–2000*, Minneapolis: University of Minnesota Press.

McLaughlin, J. (2016) 'Private intelligence firm proposes "Google" for tracking terrorists' faces', *The Intercept*, 4 November. Available from: https://theintercept.com/2016/11/04/private-intellligence-firm-proposes-google-for-tracking-terrorists-faces/ [Accessed 12 January 2024].

McLuhan, M. (1994) Understanding Media: The Extensions of Man. Cambridge, MA: The MIT Press.

Meyer, S. (2013) 'Impeding lone-wolf attacks: lessons derived from the 2011 Norway attacks', *Crime Science*, 2(7). https://doi.org/10.1186/2193-7680-2-7

Moonshot. (2023) *The Moonshot Threat Bulletin—July 2023 At a Glance*. https://moonshotteam.com/resource/the-moonshot-threat-bulletin-july-2023-at-a-glance/

Morris-Reich, A. (2016) *Race and Photography: Racial Photography as Scientific Evidence, 1876–1980*, Chicago: University of Chicago Press.

Murray, H. (2009) 'Monstrous play in negative spaces: illegible bodies and the cultural construction in biometric technology', in S. Magnet and K. Gates (eds) *The New Media of Surveillance*, New York: Routledge, pp 69–87.

Packer, J. (2010) 'What is an archive? An apparatus model for communications and media history', *Communication Review*, 13(1): 88–104. https://doi.org/10.1080/10714420903558720

Packer, J. and Reeves, J. (2020) *Killer Apps: War, Media, Machine*, Durham, NC: Duke University Press. http://dx.doi.org/10.1215/9781478007272

Packer, J. and Reeves, J. (2017) 'Taking people out: drones, media/weapons, and the coming humanectomy', in L. Parks and C. Kaplen (eds) *Life in the Age of Drones*, Durham, NC: Duke University Press, pp 261–281.

Peters, J.D. (2015) *The Marvelous Clouds: Toward a Philosophy of Elemental Media*, Chicago: University of Chicago Press.

Pick, D. (1993) *War Machine: The Rationalization of Slaughter in the Modern Age*, New Haven: Yale University Press.

Public Intelligence (2014) 'Identity dominance: the US military's biometric war in Afghanistan', *Public Intelligence*, 21 April. Available from: https://publicintelligence.net/identity-dominance/ [Accessed 12 January 2024].

Reeves, J. and Packer, J. (2013) 'Police media: the governance of territory, speed, and communication', *Communication and Critical/Cultural Studies*, 10(4): 359–384.

Reich, W (1933/1980) *The Mass Psychology of Fascism*. Trans. V. Carfagno. New York: Farrar, Straus, and Giroux.

Richards, G. (2003) *Race, Racism and Psychology: Towards a Reflexive History*, New York: Routledge.

Robertson, C. (2012) *The Passport in America: The History of a Document*, Oxford: Oxford University Press.

Robin, R.T. (2001) *The Making of the Cold War Enemy: Culture and Politics in the Military-Intellectual Complex*, Princeton: Princeton University Press.

Rorschach, H. (1942) *Psychodiagnostics*. Oxford: Grune & Stratton.

Samelson, F. (1977) 'World War I intelligence testing and the development of psychology', *History of the Behavioral Sciences*, 13(3): 274–282.

Schuurman, B. and Eijkman, Q. (2015) 'Indicators of terrorist intent and capability: tools for threat assessment', *Dynamics of Asymmetric Conflict*, 8(3): 215–231.

Schroeter, M. (2020) *Artificial Intelligence and Countering Violent Extremism: A Primer*, London: Global Network on Extremism and Technology.

Segrave, K. (2004) *Lie Detectors: A Social History*, Jefferson, NC: McFarland.

Sekula, A. (1992) 'The body and the archive in contest of meaning', in R. Bolton (ed) *The Contest of Meaning: Critical Histories of Photography*, Cambridge, MA: MIT Press, pp 343–389.

Shen, W. (2017) 'Quantitative Crisis Response', Defense Advanced Research Projects Agency. Available from: https://www.darpa.mil/program/quantitative-crisis-response [Accessed 12 January 2024].

Siegel, G. (2014) *Forensic Media: Reconstructing Accidents in Accelerated Modernity*, Durham, NC: Duke University Press.

Silva, K. (2016) *Brown Threat: Identification in the Security State*, Minneapolis: University of Minnesota Press.

Singh, J.P., Jain S., Arora, S. and Singh, U.P. (2021) 'A survey of behavioral biometric gait recognition: current success and future perspectives', *Archives of Computational Methods in Engineering*, 28: 107–148.

Sternstein, A. (2016a) 'The military is building an employee database to predict traitors', *Defense One*, 27 May. Available from: http://www.defenseone.com/technology/2016/05/military-building-employee-database-predict-traitors/128662/ [Accessed 12 January 2024].

Sternstein, A. (2016b) 'What's your "insider threat score"? It could determine if you keep your clearance', *Nextgov/FCW*, 29 April. Available from: http://www.nextgov.com/cybersecurity/2016/04/cleared-personnel-fico-score-will-decide-what-secrets-you-can-see/127903/ [Accessed 12 January 2024].

Strohm, C. (2016) 'Predicting terrorism from big data challenges US intelligence', *Bloomberg*, 13 October. Available from: https://www.bloomberg.com/news/articles/2016-10-13/predicting-terrorism-from-big-data-challenges-u-s-intelligence [Accessed 12 January 2024].

Taylor, M. (2015) 'Criminogenic qualities of the internet', *Dynamics of Asymmetric Warfare*, 8(2): 97–106.

Tucker, P. (2015) 'Why lone wolf attacks are so hard to predict', *Defense One*, 22 July. Available from: http://www.defenseone.com/business/2015/07/why-lone-wolf-attacks-are-so-hard-predict /118279/ [Accessed 12 January 2024].

Tucker, P. (2016) 'The Pentagon is developing a new suite of tools to fight the lone wolf problem', *Defense One*, 16 June. Available from: http://www.defenseone.com/technology/2016/06/pentagon-developing-new-suite-tools-fight-lone-wolf-problem/129154/ [Accessed 12 January 2024].

Tucker, W.H. (1996) *The Science and Politics of Racial Research*, Urbana: University of Illinois Press.

US Army (2020) 'Intelligence analysis', *Army Techniques Publication* 2(33.4). Available from: https://armypubs.army.mil/epubs/DR_pubs/DR_a/pdf/web/ARN20669_ATP%202-33x4%20FINAL%20WEB_v2_fix.pdf [Accessed 12 January 2024].

Zuboff, S. (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, New York: Public Affairs Publishing.

III

Politics and Ethics

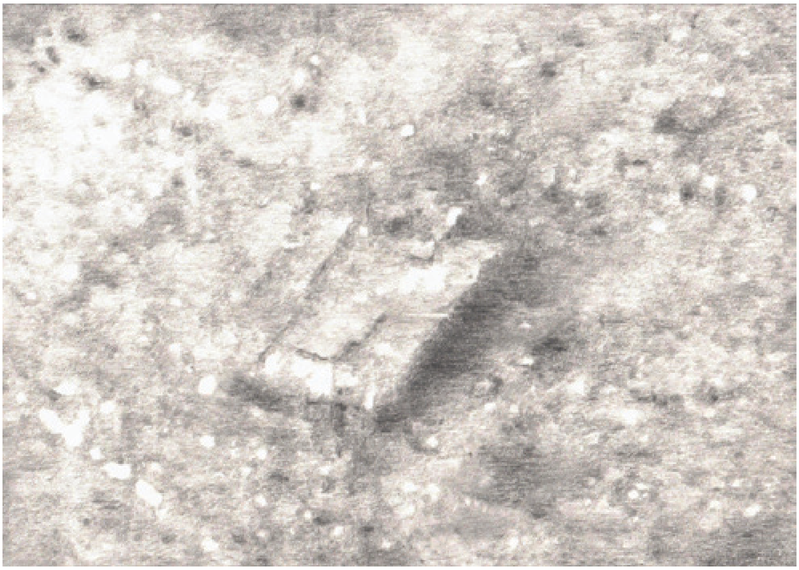
ARTWORK

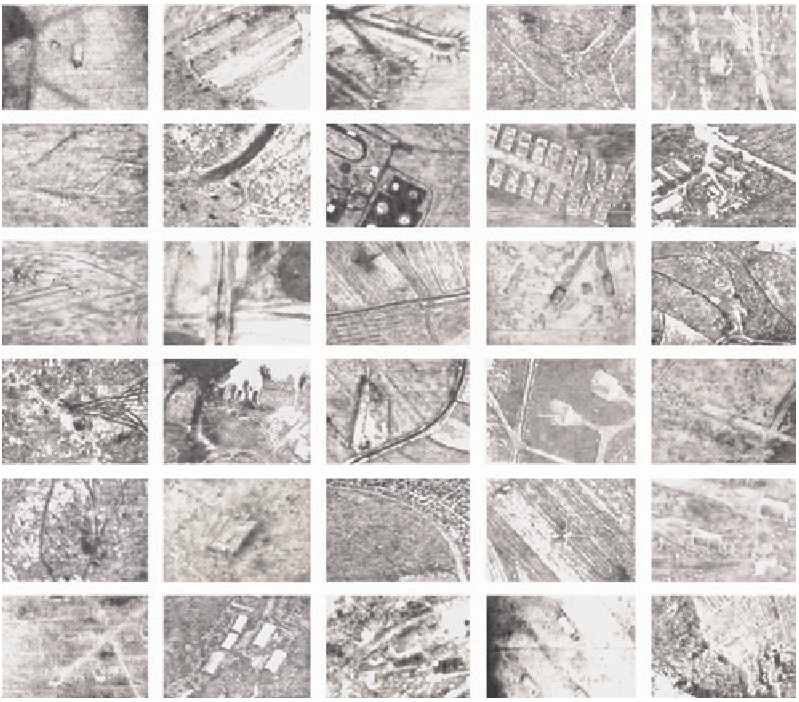
XCI|XCIX, (91|99)

Johannes Weilandt, 2023

The series XCI|XCIX, (91|99) is based on machine-generated images by precision and laser-guided weapons, so-called 'Smart Bombs', from the Second Gulf War and Yugoslav War of the 1990s, which were broadcast on television. The title of the work frames the decade that began with images of this new type of weapon (even if they rarely worked) and drove the imagination about new, technological warfare. Moreover, the images reinforced the visual culture of looking through the crosshairs. From a researched and collected selection of projectile shots on target, video stills were made in moments before hitting the target and transferred to paper by hand as a drawing in sessions lasting up to 12 hours at a time. The alienation through bodily gestural repetition in the drawing questions the disembodied images and their function to record physical erasure. At the same time, the caesura of a media reception is recalled, in which increasing immediacy produces ever more abstract images. In this way, the characteristic of the original source of the image is taken up again by the drawing: The closer the eye gets to the drawing, the more it denies its subject.

This work was nominated for the Haus am Kleistpark 2022 Award and invited to the Brecht Tage 2023 on the theme of 'Brecht's Kriegsfibel' (Brechts's War Primer) at the Literature Forum at the Brecht House in Berlin. The following images are a selection of this larger collection.





Engineering Moral Failure? The Challenges of Algorithmic Ethics for Lethal Autonomous Weapon Systems

Elke Schwarz

Introduction

The past decade of heated discussions on the ethical implications and consequences of lethal autonomous weapon systems (LAWS) has had modest effects on shaping the limits for autonomous weapons technology. Early ethical concerns raised by critics seem to have been relegated to the margins and are either superseded by largely technical concerns about how to find technology design fixes for issues of trustworthiness and control (see, for example, Roberson et al, 2022; Umbrello, 2022), or by the crafting of a set of principles – often five[1] – that guide the ethical challenges that arise with LAWS (see, for example, US Department of Defense, 2020; UK Ministry of Defence, 2022). This is a somewhat depressing state of affairs. Although the question of how to make 'responsible' artificial intelligence (AI) for the military domain is currently receiving prominent attention,[2] the solutions offered tend to focus, again, on the technical and functional

[1] The US Department of Defense has adopted a framework of five principles applicable to military AI, namely that they must be responsible, reliable, traceable, equitable and governable (US Department of Defense, 2020). Similarly, the UK Ministry of Defence has adopted the following five: responsibility, human-centricity, understandability, reliability, and bias and harm mitigation (UK Ministry of Defence, 2022).

[2] In 2023, the Dutch Ministry of Foreign Affairs, together with the Republic of Korea, hosted an inaugural global summit on Responsible AI in the Military Domain (REAIM) conference dedicated to advancing responsible AI. The outcome of this summit was a

elements of reliability, responsibility and associated lines of accountability, and not on the question of *moral* responsibility as it pertains to such systems. Yet moral responsibility is a crucial element in the ethics of warfare – traditionally so, as much as in the present. This focus on functionality and technicality is perhaps not surprising if we consider that one of the more vexing problems – that of agency, specifically human agency – becomes ever more complex and complicated the faster technological advancements proceed and the more that the human is intricately embedded in digital ecologies for military matters. A core challenge is to engage a better understanding on the role of *human* moral agency and responsibility in the development and use of autonomous and intelligent weapons systems if we wish to retain a focus on the ethics, not technics of AI-enabled weapon systems. Simply (dis)placing agency for morally relevant actions into the functional design, or indeed onto the machine, cannot be the solution for systems that decide on life and death. This much was evident early on in the conception and development of weapons systems with increasing levels of autonomy, when, perhaps, this insight was more clearly articulated than it is now.

Indeed, in the early days of AI and autonomous systems, the idea that cybernetic technologies would make a substantial contribution to the practice of warfare in ways that might prove rather challenging to human responsibility was already quite forcefully articulated by some of the pioneers of AI. Norbert Wiener, one of the key figures in the advent of automated and autonomous technology, was among the most vocal sceptics about the ostensible benefits of increased automation for, and in, warfare. He raised explicit concerns about the difficulty of having and taking responsibility for actions with increasingly autonomous weapon systems. In his seminal work on cybernetics and its relation to society, Wiener issued a clear warning, which is worth quoting in full:

> Any machine constructed for the purpose of making decisions, if it does not possess the power of learning, will be completely literal minded. Woe to us if we let it decide our conduct unless we have previously examined the laws of its actions and know fully that its conduct will be carried out on principles acceptable to us! On the other hand, the machine … which can learn and can make decisions on the basis of its learning, will in no way be obliged to make such decisions as we should have made, or should be acceptable to us. For the man who is not aware of this, for him to throw the problem of his

call to action, but little concrete advancements for a better understanding of the term ‘responsible’ has been made (see Government of the Netherlands, 2023).

> responsibility on the machine, whether it can learn or not, is to cast his responsibility to the winds and find it coming back seated on the whirlwind. (Wiener, 1954: 185)

Wiener's concerns resonate today, in a present where technologies exist that have the capacity to engage and eliminate a (human) target within a matter of seconds, and without any human agent in the actual firing-decision loop. The problem of an apparent responsibility and accountability gap that arises when machines make lethal decisions that humans have little insight into or control over is at this point very well documented (see, for example, Matthias, 2004; Sparrow, 2007; Bhuta et al, 2016), although this idea too has become contested. Various scholarship has recently proliferated that suggest such a responsibility gap does not exist, or, if it does exist, it either constitutes no problem per se or can be solved with appropriate attentiveness to technological design.[3] Some suggest that responsibility in the context of LAWS should be understood as 'distributed amongst human and non-human agents or some combination thereof' (Galliott, 2016: 228), while others argue that transferring responsibility to a machine is to be welcomed because this would mitigate the psychological impact of having to make a tragic choice in the absence of a clear moral solution (Danaher, 2022).

This type of reasoning weighs up the different elements in our understanding of human responsibility against each other. This includes the effects of responsibility for the act of killing on the responsibility holder as a significant moral factor and as an argument as to why delegating responsibility to machines might be desirable. Such discussions are increasingly finding their way into the literature on the moral utility of LAWS. This approach to ethical reasoning refracts human moral responsibility through an abstract lens, in which the concept of moral responsibility is reinterpreted to align with the technicity of causality and control. As a consequence, the place and significance of human agency shrinks, facilitated by an increased encroachment of technical-logical reasoning in matters of ethical and political thought. In this mode of ethical thinking, the complexity, irresolvability, vulnerability and plural relationality of human life as a foundation for moral reasoning takes a backseat. This also applies for military matters.

Wiener, having worked himself on many military applications of autonomous technologies, saw a twofold peril in cybernetic technologies: first, as a mode of reasoning by a group of humans 'to increase their control over the rest of the human race' (Wiener, 1954: 181); and, second, as a way of outsourcing complex and difficult decisions to machine authority 'without

---

[3] For a thorough mapping and discussion of the current debate on the responsibility gap with (L)AWS and its various positions, see Oiman (2023).

too much inquiry as to the motives and principles behind these' (Wiener, 1954: 185). Wiener was attuned to the fact that technologies are never purely technological but rather sociotechnical, in that 'scientific-technological visions enter into assemblages of materiality, meaning and morality' (Jasanoff and Kim, 2015: 4). This moral dimension is crucial but often overlooked. Every era has its form of sociotechnical ideas of military practice and, consequently, adjusting to or justifying violence. Or, as Ashis Nandy points out in 1997, when cybernetic domestic and military infrastructures had begun to sprawl, 'every age has its prototypical violence. And the prototypical violence of our age is based on secular, objective and dispassionate' pursuits of interests – or so the sociotechnical visions of the Western military discourse suggest (Nandy, 1997: 8).

As day-to-day practices, as well as military practices, are increasingly routed through digital interfaces and infrastructures – both visible and invisible – these modalities come to shape perspectives, expectations and desires in ways that prioritize a *system*atic problem-solving approach to even the most intractable of life's problems. In other words, in the contemporary era, the logic of computation and information technology has fast become a dominant mode of moral philosophical reasoning in the Anglo-American tradition, and is set to shape ethical thinking in significant ways – specifically in and for warfare.

Within such conceptions of ethics, the language of calculable utility and efficiency dominates, and the human is either objectified as a set of data points or as a functional element in the wider digital infrastructure. The literatures alluded to earlier tend to draw on a mode of abstract, technical reasoning that mirrors the technologics of the computational processes themselves. Within such frameworks, what can *moral* agency mean? What can it rest on when the task is to make the human fit for the digital environment? Thinking this through, I am reminded of the origins of the term 'cyb-org', a portmanteau coined by Clynes and Kline in 1960. The term was employed by the two researchers to suggest that rather than adapt any environment to make it fit for humans perhaps with the advent of cybernetic technology, we should strive to create 'artifact-organisms which would extend man's unconscious, self-regulatory controls', so humans could mould themselves to fit any environment (Clynes and Kline, 1960: 26). If we follow this line of thinking about humans and their situatedness, there is a tendency to prioritize the technological logic over the human in making the human 'fit' and functional for a new technological future. What kind of moral reasoning will such a perspective prioritize? To put it differently, what kind of moral reasoning do we see emerge that accommodates the sidelining of human moral agency in the use of lethal force with AI-enabled targeting systems?

The *forms* of the discussions on moral agency and responsibility take on the markers of technical discourse, even if they are held within a philosophical register. In such discussions, both human agency and machine agency are read through a technological register wherein functional equivalences are drawn between the two in order to make one fit the other. In other words, the human capacity to act is imagined in abstracted terms as purely rational, functional and technological. It is the same economic imaginary that also underpins AI, one of effectiveness and efficiency in optimizing action. This takes us further away from understanding and considering moral challenges as distinctly human and social concerns, and further into the terrain of thinking about ethics as a technological problem which could potentially be solved with more attentiveness to technology design rather than human relations.

In this chapter I engage with some recent scholarship on the moral dimensions of LAWS and examine their logical foundations. I argue that the mode of reasoning employed in such discourses is a technical mode, reflecting a desire to justify possible harm through abstract, algorithmic reasoning. In such discourses, meaningful human action and impact is all but evacuated from the discussions (O'Driscoll, 2022: 7). The upshot of this is the emergence of critical blindspots in the debate on moral responsibility for LAWS, by shifting the focus away from the impact of lethal force on humans and towards a focus on mitigating moral risks for those wielding lethal force with autonomous weapons. I begin in section 1 by briefly discussing the role AI plays for the defence sector, tracing some of the uses and imaginaries for AI in the broader weapon systems architecture. The next section then discusses recent literature on moral reasoning in favour of LAWS in which human responsibility is cast as a moral challenge to be overcome through analytic reasoning. The role of human responsibility within the broader ecology of human and digital systems is then taken on as a focus. The final section 4 problematizes the overly abstract and technical casting of human responsibility, and refocuses our attention on the social dimensions relevant to understanding what moral human responsibility entails in the context of AI-enabled weapon systems.

1. Hopes and dreams: military artificial intelligence

The debate on autonomous weapon systems and their ethical ramifications is still, to date, marred by some unhelpful conflations, imprecisions of terms or lack of agreement about what certain terms mean. This starts with the broader term 'autonomy' and stretches to concepts such as trust and responsibility. This is perhaps not surprising, given the speed with which the technology (but not the political consensus) develops, and given the transdisciplinary nature of the discussions. Trust, for example, has different layers of meaning depending on whether one considers the term from an engineering standpoint or a psychological or social standpoint. Trustworthiness as a concept in engineering primarily relates to reliability – that a system functions as it is intended to function. Trust and trustworthiness as a staple for a functioning human society are a lot more complex and multivaried (Nickel, Franssen and Kroes, 2010).

Taking seriously the concerns about conceptual conflations as hampering discussions on LAWS, I must begin with a brief delineation as to what I mean by LAWS, and the role AI plays in these systems. Definitions of autonomous systems, especially weapon systems, are not straightforward and always serve a political function.[4] The shorthand definition often used for LAWS is that such weapon systems have autonomy in their critical functions – select, track and attack targets without human intervention in the action loop (International Committee of the Red Cross, 2016). Weapon systems with limited autonomy that employ sensors for the task of selecting, tracking and attacking incoming threats have been around since the late 1970s.[5] These systems, while technically autonomous in their critical functions, are not my concern here; my concern is with those systems that employ advanced AI techniques to execute critical functions, specifically those that employ AI techniques to algorithmically identify what should be designated as a target and make a lethal decision based on this algorithmic target identification.

However, discourses on AI are also fraught with differing ideas, visions and understandings of just what AI is. I follow Lucy Suchman (2023) in drawing a distinction between AI as a technique and technology, and AI as a signifier which works to serve the interests of those giving meaning to the sign. The former seeks to demystify the term 'AI' and focuses on the operational dimensions of the term. With this in mind, AI is a computational technique to 'extract statistical correlations (designated as patterns) from large datasets, based on the adjustment of relevant parameters according to either internally or externally generated feedback' (Suchman, 2023: 2). In order to facilitate these operations, certain techniques to gather, code and classify data are required. AI is thus not a 'thing' or singular 'object', but rather a set of techniques and technologies.

Yet, in current military and industry imaginaries, AI represents a lot more than this. Here, Suchman's reading of AI as a floating signifier in current discourses on AI is on point. She suggests that AI is 'a term that suggests a specific referent but works to escape definition in order to maximize its suggestive power' (Suchman, 2023: 3). At present, this suggestive power is

4 For a detailed discussion on the challenges of terminology in the international debate on autonomous weapons systems, see Bächle and Bareis (2022).

5 One of the first such system was the CAPTOR mine, which used sound propagation to detect, track and launch its weapon component against hostile submarines (Trouver, 2012). Modern versions of such a limited autonomy system that can sense and react independent of human action, such as the C-Ram, the Phalanx or the Mantis System, have been in use for many years (Bächle and Bareis, 2022: 13).

tied up with a persistent anthropomorphism by which the AI system is cast as an intelligent objective and neutral agent with the capacities to overcome human cognitive limitations. The early iterations of AI as 'thinking machine' paved the way for this anthropomorphized imaginary of AI as agent.

For quite some time now, AI has been the buzzword not just for business ventures but also for military organizations across the globe. It is both incredibly attractive as a promise and incredibly lucrative as a product. The global military AI market is growing quickly – it was estimated to be worth just under US$9.2 billion in 2023 and is projected to at least double in the next five years (*Markets and Markets*, 2023). This, in turn, drives growth in the autonomous weapons industry, a market that is forecast to grow from an estimated US$13.3 billion in 2022 to US$21.8 billion by 2027 (Business Research Company, 2023). The US and China are currently the dominant drivers of this market, ahead of other states like the UK, Israel, France and Russia in accelerating military AI rollout.[6]

Military AI can take many forms and serve many purposes that benefit from speed and optimization. AI technologies can, for example, be useful for logistics operations, locating assets more efficiently or getting equipment from point A to point B; they can also be useful for predictive maintenance of vehicles and equipment, or to streamline infrastructural processes. This is what AI can do best – to increase the speed and efficiency of narrowly set tasks within a controllable environment with limited parameters. Such areas of use are, with some caveats, relatively uncontroversial. AI for narrow, limited tasks is likely to work most reliably and with predictable benefits for military operations in the near future (and likely beyond).

But there are much grander and more expansive visions in play for military AI. In such imaginaries, the role of AI becomes much more pervasive in order to facilitate action across distant geographies at high speed. The US Joint All Domain Command and Control (JADC2) programme is illustrative of this aim. The JADC2 vision is one which

[6] China's role in the broader global development of military AI and autonomous weapons systems is significant, but it is not the subject of this chapter. For the purposes of the argument I seek to make here, I focus on the Anglo-American context and US discourses on military AI and LAWS. I do so because the US is among the most vocal, transparent and explicit in both its aspirations for the ubiquity of AI and the ethical dimensions it considers for its military AI. Moreover, the US has a longer active history of the use of digital technologies in active conflicts and has shown itself to be a first mover in the rollout and use of digitally networked platforms for target identification through programmes such as Project Maven. Moreover, the US is a key producer of military AI that will be used, in the interests of interoperability, by other allied states (Schmidt et al, 2021: 83). Finally, the discourses this chapter is concerned with build on the history of US development of cybernetic technologies and therefore should be analysed in this broader historical sociotechno-political context.

strives for 'shared situational awareness, synchronous and asynchronous global collaboration, strategic and operational joint planning, real-time global force visualization and management, predictive force readiness and logistics, real-time synchronization and integration of kinetic and non-kinetic joint and long-range precision fires' (US Department of Defense, 2022) – an imaginary of technological omniscience and omnipotence, or, at the very least, 'decision dominance' (Lloyd and Rozman, 2022) that stretches, in theory, across the globe. More recently, the advent of large language models such as OpenAI's ChatGPT has opened up visions for AI-driven platforms to assist in all aspects of battle decision making. The US-based company Palantir, for example, is marketing its large language model-based AI platform (AIP) to militaries across the globe. According to Palantir, AIP connects 'highly sensitive and classified intelligence data to create a real-time representation of your environment' (Palantir, 2023). It is, in essence, a chatbot to help 'order drone reconnaissance, generate several plans of attack and organise the jamming of enemy communication' (Gault, 2023).

The aim is to integrate AI into most if not all functions of military operations, including the targeting cycle, where the ethical and legal stakes are high. The report by the US National Security Commission on AI is instructive as to what the goals for military AI are in the near future. The 2021 Commission report clearly recommends as follows: 'The Department [of Defense] must act now to integrate AI into critical functions, existing systems, exercises and wargames to become an AI-ready force by 2025' (Schmidt et al, 2021: 77). The fact that the US government 'still operates at human speed, not machine speed' is seen as a considerable detriment to progress and a disadvantage that must be overcome. Former Google CEO Eric Schmidt, the chair of the National Security Commission on Artificial Intelligence (NSCAI) and its report, has been an outspoken proponent of making a fully AI-ready US military organization. To do this, he envisions an 'internet of things with a deadly twist' (Knight, 2023).

In an AI-facilitated accelerated target identification and attack cycle, the role of human decision making throughout this action sequence, and in particular at the sharp point of delivery, becomes necessarily marginal. This is not an accident, but ultimately the stated desired goal by those advocating for more military AI, who see human cognition as insufficient in future wars (Schmidt et al, 2021: 25, 90–98). Already, since 2018, we have seen the goalposts shift considerably in terms of where a human can or should exert a control function for systems that employ AI for targeting actions. The early iterations of what meaningful or appropriate human control constitutes tended to insist on the human as a final arbiter in the lethal decision – in other words, that control always (and meaningfully) rests with a human pulling the trigger. This has shifted to an interpretation that meaningful control can or should indeed be distributed and can be exerted appropriately at the design, parameter input and/or supervision stage – a system of control of sorts (McFarland, 2022). In all this, the shift towards human-AI teaming becomes dominant whereby the AI is considered a 'true partner' to help cut through the fog of war, as the Defense Advanced Research Projects Agency's (DARPA) Project Squad X, for example, aspires to. Many such ventures of human-machine teaming are underway to assist in this effort towards a complete human-machine merger (DARPA, 2019). For this purpose, BAE Systems, via Lockheed Martin, is developing AI to 'inform human decision making in complex, time-critical combat situations' for an increased battlespace and sphere of impact (BAE Systems, 2020). In other words, the technology is intended to provide the on-the-ground human squad with better and faster intelligence to act on for a range of aspects, including targeting. At the time of writing, in 2024, such visions are aspirational. There is no clear evidence that AI can indeed lift the proverbial or literal fog of war in highly dynamic, complex and uncertain contexts of conflict (Walsh, 2022). On the contrary, some argue that in addition to the fog of war, such systems will likely exacerbate uncertainty through 'systems fog' (Hughes, 2020). At this stage, and likely in the not-too-distant future and beyond, the level of sophistication in AI weapons systems is no match for the multivaried quirks and complexities of the real world (Walker, 2021). Nonetheless, the pursuit to draw in ever more AI-enabled machine decision making and place the human ever more intricately at the functional margins of the lethal system continues unimpeded.

These recent advancements in the human-AI teaming domain have given greater urgency to the need to get to grips with the complex ethics that such systems raise, not least because the speed with which military technology start-ups churn out military AI products has accelerated the possibility for harmful autonomy in lethal targeting (Hoijtink, 2022; Holland Michel, 2023: 22–25). So what then of human agency and moral responsibility when the human cognitive functions are insufficient and need to be supplemented, if not replaced by machine decisions? How can an operator or commander still exercise moral judgement in an accelerated AI-enabled kill chain? As I have argued elsewhere, it is not possible to viably and meaningfully exercise moral responsibility in such a configuration (Schwarz, 2021). Nonetheless, for some, the answer to the pressing question of responsibility and agency is that perhaps it does not matter so much whether the human retains responsibility and agency, indeed, that such agency is undesirable, and better placed within the machine, justified either through 'ethics-by-engineering', or a moral wager. I will now briefly turn my attention to these accounts.

2. Responsibility by calculus or by design

Understanding and accepting the ethical challenges of AI-enabled autonomous weapons is fraught with difficulties. The ethics of any weapons system is a matter of debate, but specifically systems that take on functions which otherwise are ascribed to human actors and that, in doing so, are by some considered as 'more ethical' than humans require careful and ongoing ethical review and consideration. An early proponent of LAWS as a possible remedy to human atrocities in warfare was the roboticist Ronald Arkin, who proposed that autonomous weapon systems might, in the future, be able to perform better than humans in the heat of the battle or the fog of war because robotic systems are not marred by all the human flaws that come with warfighting – emotions, psychological drives, cognitive limitations, a tendency to overreach or overreact and the always latent possibility of erratic or unpredictable behaviour in battlefield scenarios (Arkin, 2009: 6–7). Autonomous robotic systems, he argues, could instead act as better ethical agents as they might be equipped with an ethics module, in which the laws of war and the rule of engagements are inscribed and according to which lawful decisions can be made. Since then, others have taken up the task to weigh the fallibility of humans in war against the promise of a smooth functioning technological agent (see also Heller, 2023).

A recent set of literature in the analytic philosophy vein, for example, develops the case in the interest of the protection of humans, namely that, on the whole, it is desirable, if not morally mandated, to develop and use LAWS if they mitigate risk to the human in warfare. In 2022, for example, Eric Riesen argued that a moral case in favour of LAWS can be made on the grounds that they would protect soldiers not only from the risk of physical harm, but also from psychological harm and, importantly, moral risks (Riesen, 2022). This lowered risk to one's own combatants through LAWS constitutes, for Riesen, 'massive moral gains made possible by the technology' (Riesen, 2022: 133) and should therefore become morally mandated rather than banned. To argue his case, Riesen employs ethical reasoning in the mode of 'if/then' algorithmic logic:

> If X gives Y an order to accomplish good goal G, then X has an obligation, other things being equal, to choose a means to accomplish G that does not violate the demands of justice, make the world worse or expose Y to potential lethal, moral or psychological risk unless incurring such risk aids in the accomplishment of G in some way that cannot be gained via less risky means. (Riesen, 2022: 136)

Bracketing, for the moment, the indeterminacy of what it might mean to 'not violate the demands of justice' or how to measure what might 'make the world worse', it is easy to see how this moral mandate to safeguard the physical, moral and psychological lives of service personnel is clearly prioritized in this calculative formula. In Riesen's view, this promise for moral progress also outweighs most countervailing concerns, including those one might have about a gap in responsibility for unwarranted harms caused through autonomous actions, as highlighted by Sparrow and others. Indeed, this too, Riesen argues, can be addressed with a sufficiently technological perspective. To do so, he draws on the work of Thomas Simpson and Vincent Müller (2015), who, in turn, borrow 'the concept of a "tolerance level" from engineering ethics' (Riesen, 2022: 142). Setting tolerance levels means artificially setting upper and lower limits within which systems are expected to function as intended. By setting such limits, the range of use and expected functioning of any engineered system – and within this, the most expected outcomes – could be determined. For these actions, either a regulator or the systems engineer could be held responsible. If systems are used outside of the set parameters and something goes wrong, the user is responsible. And for anything unforeseen, a residue gap in blameworthiness is to be expected but acceptable, as 'the strength of the positive case outweighs the badness of the remaining blameworthiness gap'. The bottom line here is that the moral burden of killing should not unnecessarily rest on the human (Riesen, 2022: 142). In that sense, responsibility is a moral risk.

Cappuccio et al (2022) have argued similarly, against the desirability of human responsibility in the act of killing and state that it is morally commendable, if not imperative, to employ LAWS, as they would help mitigate moral injury. Here too, the analysis is framed by mathematical reasoning, built on two premises ('only entities with a moral conscience risk to suffer trauma when given an order to kill; and [m]inimising warfare-related trauma is ethically imperative'), which, *if* true, it must *then* follow that 'it is ethically imperative to replace human soldiers with autonomous weapon systems … capable to independently search for and engage enemy targets' (Cappuccio et al, 2022: 2). Foregrounded here is not the harm inflicted on possible human targets; instead, the psychological and mental risk that comes with taking another human life in warfare is seen as an ethical challenge to overcome. In other words, the problem is not the killing, but rather what the killing does to those charged with killing. This seems like a rather peculiar reorientation of moral concerns, but it is made possible by a mode of logically consistent calculative reasoning that works with variables and weightings, which tend to be assigned based on the moral philosopher's intuitive priorities, which, in turn, remain underspecified. Here the moral weight of carrying the responsibility for the act of killing, and its potential psychological implications, is seen as the key moral variable to account for.

Examples like these follow in the tracks of a long lineage of analytic hyper-rational reasoning in which costs and benefits are weighted against each other in the justification of harms in the context of war. Often, this technical mode of reasoning is done with the help of somewhat outlandish hypothetical case scenarios in which moral agents are thrust into all manner of emergency decision scenarios in which the moral utility of decision A is weighed up against that of decision B, as put forward, for example, by the famous Trolley problem[7] (Schwarz, 2024). These approaches purport to employ rigorous abstract reasoning and arrive at their moral truths through 'detachment, abstraction, universalism', in the mode of scientific inquiry and technological possibility that places a premium on 'analytic rigour and argumentative precision' (O'Driscoll, 2022: 7). However, such modes of ethical reasoning are only coherent if ethics is understood to work 'like a science' in that the subject to be assessed morally is neatly categorized into relevant variables which can be assessed through algorithmic logics and scientific hypothesizing (McMahan, 2013; Schwarz, 2018: 131). Yet, with this technical approach, this type of discourse about the ethics of war, and its associated weapons, becomes, as O'Driscoll (2022: 7) notes, blind to its own subjective position in favour of an assumed objectivity. Nevertheless, it is precisely this type of reasoning that seems to lead to an 'evacuation of the human element' from thinking ethically in and about war (O'Driscoll, 2022: 7).

3. Displacing the human from responsibility

The logic of the technology and the logic of this mode of ethical reasoning, then, go hand in hand in whisking the human away from his or her moral responsibility, as the human is cast as a vulnerability and liability, not just in the accelerated mode of warfare possible with LAWS, but also within its frames of moral responsibility. Nonetheless, the question of responsibility persists in almost all discourses on the ethics of LAWS. But now, it becomes a question of shifting the locus of responsibility, away from human actors, towards a more complex picture of distributed

[7] The Trolley problem was originally put forward by the Oxford philosopher Philippa Foot, who posited a decision scenario in which a runaway tram barrels down a track on which there are five people who cannot leave. The tram's brakes fail and the conductor cannot stop the tram, but he can divert the runaway carriage onto another track where only one person is trapped. The question raised is whether it is morally permissible to divert the tram to kill one instead of five. Later iterations of this problem saw the tram become a trolley and the choice to divert was made by a bystander who could flip a switch (Foot, 1967). Foot's original experiment spawned an entire industry of hypothetical case scenarios (Edmonds, 2015) and the mode of reasoning has found its way into contemporary discussions on the ethics of self-driving cars (see, for example, the Moral Machine experiment: Awad et al, 2018).

or, indeed, entirely relocated responsibility. The concept of a distributed responsibility as a virtue – meaning that responsibility is allocated to both the human and the machine as an agent – suggests, as Duncan MacIntosh explains, that a commander should not really have to carry the burden of having to make a difficult lethal decision alone. Instead, it should be possible to distribute 'the cost of conscience' and if the distribution of the difficult decision as to 'whom exactly to kill and when' can be 'delegated to a morally discerning but morally conscienceless machine, we have the additional virtue that the moral offloading … is done onto a device that will not be morally harmed by the decision it must make' (MacIntosh, 2021: 14). Sparing the conscience of a moral agent is thus prioritized in MacIntosh's moral calculation. For him, the decision to outsource the decision to kill, and with that the moral responsibility of doing so, to the machine is, in some cases, 'morally less bad' because machines 'can randomize and democratize violence' (MacIntosh, 2021: 16). The moral gravity of the act of killing is acknowledged, but not as a prompt to moral restraint, but rather as an obstacle that can (and should) be overcome with more responsibility bestowed upon the machine.[8]

Jai Galliott also makes the case for distributed responsibility, but one that stretched across various human and nonhuman parts of the human-machine assembly. Galliott suggests that everyone involved in the machine system – including the programmer, engineer, manufacturer and/or commanding officer – all have some degree of moral responsibility to bear for actions that are morally relevant and that are executed by autonomous systems. But rather than focus on individual responsibility, some sense of collective responsibility is more appropriate for Galliott, not least because this 'has the means and scope to include non-human action' (Galliott, 2020: 171). In doing so, Galliott embraces a pragmatic approach which suggests expanding the category of moral agency to become 'a continuum of agency between nonmoral and fully moral agents, with the sort of robots we are concerned with here falling just short of the latter' (Galliott, 2020: 171). With this, Galliott is right to see the moral implications of sociotechnical systems as one of complexity. But, I would suggest, there is no equivalency between accepting responsibility for a poorly engineered algorithm or a shoddily constructed computer module which results in inaccurate inventory data, for example, and accepting responsibility for taking a human life. The idea of collective responsibility, or indeed collective guilt is fraught with complexity and Hannah Arendt, for example, has argued that 'where all

[8] Others have made similar arguments recently about the psychological burden of moral responsibility in tragic-moral choice scenarios and advocate *for* a techno-responsibility gap, albeit not specifically in the context of LAWS. See, for example, Danaher, 2022; Munch et al, 2023.

are guilty, nobody is' (Arendt, 2003: 147).[9] Nonetheless, Galliott endorses taking 'a pragmatic (or functional) approach', whereby 'moral responsibility [becomes] more a matter of rational and socially efficient policy that is largely outcomes-focused' (Galliott, 2020: 171).

Casting moral responsibility in terms of distributed responsibility which can be assigned along various points by design (either systems design or practice design) is finding its way into literatures that emphasize the possibility to employ value sensitive design (VSD) for meaningful human control.[10] Echoing both, the possibility Arkin proposed of building a robotic system that adheres to certain prestipulated values, and the perceived need to overcome the responsibility gap issue, this literature too tends to see the issue from a more practical and increasingly technical standpoint. The priority here is to get to grips with the moral problem of potentially unattributable moral responsibility with LAWS through systematic problem-solving approaches whereby moral responsibility becomes a functional category. In other words, a possibly intractable problem of a gap in moral responsibility, or the problem of a lack of *meaningful* human control, is sought to be solved through various technical approaches and fixes to the general categories of 'responsibility' or 'control'.

In doing so, the temptation to conflate functional autonomy, in the engineering sense, with human autonomy as understood in philosophy appears to be strong in the context of LAWS. The boundaries between what constitutes machine functionality and what constitutes moral agency become increasingly fuzzy. It is perhaps not surprising when AI is conceived as more than a very potent statistical data processing technology but as a teammate which the human can and should trust in the heat of battle, as indicated earlier. I am not suggesting that the scholars I have discussed here are needlessly anthropomorphizing the technology;[11]

[9] The perils of this for technologically distributed responsibility are illustrated in the case of an autonomous vehicle crash that killed a female pedestrian, Elaine Herzberger. Multiple factors played a role in causing the fatal accident, including the system's engineered break delay to facilitate driver comfort, the fact that the 'system design did not include a consideration for jaywalking pedestrians' (Shepardson, 2019) and the driver's lack of attentiveness to what was going on while the autonomous mode was activated. It was the human driver who was ultimately legally charged with reckless endangerment, not vehicular homicide. The manufacturer's stipulations were that it is the driver's responsibility to keep an eye on the road, even though decades of cognitive scholarship suggests that this is near-impossible as a task (see, for example, Cummings, 2004). The autonomous car company, Uber, was acquitted of all (legal) responsibility. Ultimately, nobody ended up taking responsibility for Herzberger's death (Smiley, 2023).

[10] See, for example, Ekelhof, 2019; Verdiesen, 2019; and Umbrello, 2022; among others.

[11] Although it is perhaps worth pointing out that many of the texts do depart from an understanding of the technology as ideal-type technology, which functions as desired

rather, I am suggesting that the temptation to assess and analyse as-of-yet emerging technologies and our human relationship with these new weapon systems in predominantly functional and technical terms appears to be increasingly dominant. Perhaps that is not terribly surprising in the complex relations between technologies that were conceived of as 'thinking machines' (Turing, 1950), and where communications is viewed as being twinned with control (Wiener, 1954: 16–18). Unlike other technological innovation, cybernetics, as Hans Jonas had pointed out in 1966, has brought forth a peculiar condition: 'there is a strong and, it seems, almost irresistible tendency in the human mind to interpret human functions in terms of the artifacts that take their place, and artifacts in terms of the replaced human functions' (2001: 110). While this is a condition that, to some degree, applies to most technological innovations in some form, what is different with cybernetic technologies is that it is through the advent of this type of technology that we begin to read 'human behaviour, processes of thought and sociocultural organisms' through the logics of cybernetics. In other words, it facilitated thinking about not just the material but also the mental phenomenon in the same functional register (Jonas, 2001: 110). With this lens, the human and the machine become, functionally, one and the same, both subject to the demands of functionality, perfectibility, improvement and problem solving. As I explain elsewhere, Günther Anders has identified this condition as *Promethean Shame* (Anders, 2010; Schwarz, 2019), whereby the human measures himself or herself against the standards of perfectible machines. Similarly, Hannah Arendt observed in 1970 that cybernetic ideas were seeping into the social and political fabric as a primary mode of dealing with the world's complexities (1970: 6–7). And even Norbert Wiener himself was aware that something significant was underway in shaping the ways in which humans think about themselves and their affairs, in realizing that cybernetic modes had become a type of 'applied social and moral philosophy' (Heims, 1989: xii). With this, something of these human affairs is rerouted, if not lost. Wiener writes that:

> When human atoms are knit into an organisation in which they are used, not in their full right as responsible human beings, but as cogs and leavers and rods, it matters little that their raw material is flesh and blood. *What is used as an element in a machine is, in fact an element in the machine.* (1954: 185, emphasis in original)

and without significant difficulties, errors or failures. However, to date, this assumption cannot be supported by evidence.

Wiener did not have a specific theoretical understanding of moral agency, or indeed moral responsibility, but it did not escape his notice that there is a relationship between embodied human beings worth preserving.

It is useful to return to these early thinkers in considering how the contemporary lens, in which moral responsibility as a functional element is foregrounded, has been forged. They, perhaps, help us see something we are no longer able to see in our present technoscientific condition, namely that there is a gradual erasure of the human as the sole focus of intractable moral challenges. There are, perhaps, psycho-pathologies that help explain this voluntary erasure of embodied human relationality, especially in the context of warfare. Ashis Nandy, for example, suggests that the turning to a scientific-technological mode is an upshot of trauma incurred in the First World War, which had such physically devastating effects that it required a certain 'ego defence' in which cognition and affect are split to help 'the human mind cope with unacceptable or ego-alien impulses and external threats' (Nandy, 1997: 8). Seeking to make sense of the mindless, large-scale bloodshed of the First World War raised pressing questions about the human ability to engage in violent pursuits without being 'burdened by feelings' (Nandy, 1997: 8), the assumption being that humans are not naturally inclined to kill other humans without emotional distress. For Nandy, the ability to do so is a clear upshot of the modern, industrialized condition. Following Freud, he suggests that the mechanism of isolating an unpalatable, unpleasant or unacceptable event – such as killing a fellow human – from one's significant and formative experience is to 'cauterize it emotionally'. In other words, splitting rationality and affect serves as a psychological defence mechanism to justify rationally what emotionally would otherwise be unjustifiable. This is both the upshot of prioritizing objective scientific thinking and a potentially dangerous pathology to which we owe some of the most abhorrent practices of violence in recent history. In the wake of the Second World War, psychoanalysis sought to make sense of the inhumane acts committed under fascist rule – as a 'typical psychopathology of the modern world … namely the ability to partition away human cognition and pursue this cognition unbridled by emotional or moral constraints' (Nandy, 1997: 9). Under the mandate of scientific rationality and mechanistic systems, objectivity gives way to objectification, processes become routinized and 'it becomes easy to forget the nature of the product that emerges from this process' (Kelman, 1973: 47).

This is not the place to further explore these psychoanalytical trajectories, but highlighting them helps us understand that the human relationship with other humans is fraught with challenges that cannot in all instances be technologically solved. On the contrary, a technological approach serves to further push the psychopathology of emotional detachment for rational and possibly atrocious acts into the realms of normality.

The delegation of moral responsibility to machines, and our blindness to doing so, comes at a cost. Christopher Coker rightly reminds us that precisely this was the essential point of Karel Čapek's play. 'The moral of Čapek's fable', Coker writes, 'is that we are human only to the extent that others recognise their humanity in us' (2013: 19–20). So, what, then, is moral responsibility when the human is no more than a functional variable and when the ethical act has no clear human referent? Perhaps we now need to have some clarity on the concept of moral agency as a distinct category of agency.

4. The social fabric of moral agency

A standard starting point from which ideas of what it takes for an agent to take moral responsibility is that offered by John M. Fischer and Mark Ravizza (1998): moral responsibility begins with the premise that humans are unique in that only they can be held responsible for their actions, unlike other creatures or entities. For an individual to take moral responsibility, he or she must see themselves as a *moral* agent and must accept that they are a 'fair target of the reactive attitudes' (Fischer and Ravizza, 1998: 211) in response to their actions, based on the condition of being able to act freely (that is, not under duress or forced) and within reasonable assumptions of knowledge (which guides foreseeability of outcomes). As such, they must be able to make a decision *freely* and *knowingly*. The focus in such an account of moral agency is primarily on the individual rational agent, as it is tradition for Anglo-American reasoning, but what it also implicitly acknowledges is that this agent and their moral understanding is socially situated. In other words, it is relational and dependent on the existence and views of other humans. Thus, *moral* agency, different from agency as such, is always a socially relational attribute that has meaning only among and between humans and cannot be untethered from this context. As humans, we understand and are able to judge the specificities of human relations and relationality in a range of social contexts in a way that technological artefacts are simply not able to. Or, to put it another way, our morality (and thus our moral decision making) is anchored in our history of human social relations. It is this condition that makes us not just actors, but also moral actors. Moral agency has both an individual basis and a social basis.

Human moral agency in relation to AI technologies is far from being a settled issue. In trying to make sense of this relationship, it is tempting to find analogies: Some suggest that the interaction between human and AI-enabled weapon might be comparable to that we might have with a service dog, or with a child (soldier), or with mercenaries, whereby similar structures of command, control and responsibility might apply. One should, as Deane Baker, for example, argues, consider if, and how, LAWS are different from weaponized animals or private military contractors in thinking through the ethical or legal objections to LAWS (Baker, 2022). However, it is more complicated than that and so we must reckon with what happens to human moral agency when the human is embedded in a digital information structure. And there are competing accounts as to what might constitute moral agency in a digitally mediated environment.[12] When considering *moral* agency in the context of autonomous systems, it is worth making a clear distinction between human agency and artificial agency. AI is a product, an artefact and, unlike humans, not a 'natural' or sentient being. Its anthropomorphic qualities might give humans a sense of shared meaning and understanding, but AI as a technique and technology is incapable of understanding – which is at the core of moral meaning making. While some theoretical approaches rightly decentre the human as the sole agent in the wider network of planetary life, and specifically within digital networks (Amoore, 2020), the question remains as to what degree ethics and moral agency can have meaning outside the human realm (Schwarz, 2017). Hence, I focus here solely on the human as the bearer of meaning for moral acts. However, this, too, is not straightforward. One soon realizes that there are different and contested ideas, varying by cultural context, geography or realm of inquiry. In Anglo-American discourses on moral agency alone, there are differing ideas as to the degrees of autonomy that a moral agent is perceived to have.

There is, for example, a long and ongoing debate between those who consider free will to be the driving force in the face of discourses that suggest that we are all but determined by internal or external drives (Kane, 2002). There are those who hold that free will ultimately does not exist, but is an illusion. And there are those who are finding their position somewhere in the middle: the compatibilists.[13] The content of this particular debate shall not detain us for the moment; rather, it is noteworthy that foregrounded in such conceptions is the individual as a rational moral actor. Towards whom the action is directed often plays little or no role; instead, both the harmful action itself and the recipient of the action are cast in abstracted and detached terms. As we saw in the preceding examples, the moral challenge is not the death of another human, whoever that human might be, but rather the risk incurred to the moral actor. The horror of killing is only significant insofar as it constitutes a possible harm to the actor.

This focus on the individual and the construction of moral theorizing around the right or wrong individual actions is not irrelevant. Nonetheless,

[12] See, for example, Sullins, 2006; Himma, 2009; Bryson, 2010; Gunkel, 2017; and Gellers, 2020. to name but a few.

[13] For a comprehensive discussion of these positions and their respective proponents, see Fischer et al (2007).

it betrays an unduly rationalist and unfeasibly detached approach to thinking about moral agency. Morality and our actions in relation to moral choices are not simply a system of rules that can be employed, although it is broadly recognized that:

> many philosophers think so and regard ethics as the discipline that formulates, systematizes and justifies such rules … The culprit is almost always, an excessive rationalism which takes morality to be an abstract system of principles whose truth no fully rational soul who gave them a complete and impartial hearing could deny. [But] what this approach obscures and distorts when it runs to such abstractions is morality's social function, its role in defining and regulating people's personal relation and their more distant social relations. (Deigh, 1996: 1)

There is an intricate and important relationship between moral agency and trust. It is through trust that mutual acceptances of moral allowances and prohibitions are established and it is on this trust that our actions and expected reactive attitudes for our actions rest. It is because we understand ourselves, as humans, in relation to other, embodied, sensing, feeling beings, who are affected by our acts, that expectations about our moral acts are shaped. Most importantly, it is a recognition of mutual vulnerability. A central component to human society, trust has its foundations in the shared human condition of vulnerability and dependence on others (Ess, 2020: 405). This is precisely the point: as humans we know and implicitly acknowledge that we are mutually vulnerable, for better or for worse. This is the grounds for a moral demand vis-à-vis another being. After all, 'what can responsibility mean without the risk of exposure to chance, without vulnerability' (Keenen, 1997: 51).

Moral agency, then, is both relational and affective. This social connection is important to keep in mind when considering the question of moral responsibility for harmful actions, especially within networks and assemblages of multiple human and nonhuman parts wherein the harmful action is decided by the nonhuman. The origins of the term 'responsibility' are composed of the Latin terms *re* (again) and *spondere* (to pledge). Implied in this term is, in other words, an obligation and expectation, a promise to pledge something to someone in return. Within this interpersonal relationship resides the moral dimension of agency. However, this interpersonal grounding is not static, but rather dynamic and requires reflexivity about oneself and one's embeddedness in multiple social contexts.

Here, Alasdair MacIntyre's (1999) insights about the complexities of moral agency for anyone embedded in modern bureaucratic structures are relevant and useful if we wish to contextualize the rational individual moral actor. He too stresses the core condition of morality, namely that morality and moral agency are not simply about following established and accepted rules, but rather about reflecting on these rules and their impact socially. This is particularly vital in a fragmented, highly bureaucratized modern setting in which moral agents inhabit multiple roles. For MacIntyre, moral agency means engaging with the tensions that might arise in being placed in a setting with competing and irresolvable rules about what is right and what is wrong. It is precisely this that makes an agent a moral agent. The more restricted and bureaucratized the roles, the more difficult it is to exercise this moral agency, but the divided self nonetheless has a moral responsibility to grapple with competing moral claims in such settings, this does not absolve this divided self from responsibility. As he says:

> They may inhabit a type of social and cultural order whose structures to some large degree inhibit the exercise of the powers of moral agency. But they share in responsibility for having made themselves into the kind of diminished agent that they are. Their responsibility is that of co-conspirators, engaged together in a conspiracy that functions so that they can lead blamelessly compliant lives, able to plead lack of knowledge as well as lack of control over outcome for which they might otherwise be jointly responsible. (MacIntrye, 1999: 327)

This is particularly relevant for structures that are informed by cybernetics logics. As highlighted by Jonas, Anders, Wiener and Arendt, the risks to moral agency are, as MacIntyre also notes here, that the digitally embedded moral actors' lives 'express the social and cultural order they inhabit in such a way that they have become unable to recognise, let alone to transcend its limitations' (MacIntyre, 1999: 327). We are thus left with a myopic, rationalist, unidirectional understanding of moral responsibility.

Conclusion

Our moral and cognitive worlds are tightly intertwined; reconfiguring a social and moral world requires reordering a cognitive one and vice versa. The imaginaries we inadvertently draw on for this reordering matter. We should be mindful not to reorder our moral world in the image of a rational, smooth-functioning but ultimately morally limited machine logic. Technological rationales have social upshots as the two are entwined. Considering LAWS as sociotechnical weapon systems is important, but there is perhaps a risk that our understanding of the social dimension in this becomes shaped and guided by technological ideas. Taking the social and the technological domains as intrinsically interlaced helps us better understand the interplay between technology and human agency and the co-constitutive nature in what happens when technological artefacts and human lives (singular as well as plural) come into contact. Such a focus can highlight the importance of recognizing how the social is shaped by external dynamics, including technological dynamics. It can also extend in the opposite direction and attention might be focused on how to make the social pliable for the technical. What this chapter has sought to highlight is precisely this dynamic by which the social is routed through technological imaginaries.

What sneaks into this mode of thinking about technology and its ethical ramifications is a progressive alignment with machine learning logic as an 'abstract utilitarianism, a mode of calculative ordering that results in particular ways of structuring systems' and thought (McQuillan, 2022: 13). Especially if AI is cast as a Promethean actor in its own right of sorts, rather than a set of techniques for statistical data processing, it is easy to see how scientific-technological imaginaries enter into assemblages of meaning and morality, and complicate notions of control or responsibility.

A thorough analysis and comprehension of moral agency and moral responsibility is required if we wish to better understand why this role cannot viably be inhabited by nonhuman elements, or indeed shared by AI teammates. No system has the capacity for such relational, affective and reflexive socially embedded reasoning. This should matter deeply for decisions of life and death. The AI system can at best mimic such a foundation, but rationally and affectively, we might always know that there are discrepancies. Ultimately, the algorithmic decision is not the same type of decision as a human, or humans, would make it. When we 'weigh up' morally relevant decisions, we do not simply apply a formula or entirely rationally calculate one outcome against another. It is not, as Thomas Beschorner and Florian Krause (2021) point out, a mechanical weighing similar to mathematical optimization, unlike the coded decision path. Human decision making has what they call a 'fictional surplus' (Beschorner and Krause, 2021: 81), which means it includes imaginaries about the future and about possible outcomes in uncertain conditions. This fictional surplus relies on the history of human relations and experience. Moreover, our human decision-making process is not linear, and nor does it adhere to the linear steps in an algorithm. It is possible for us to make unexpected decisions for which one then is expected to take responsibility. Finally, and importantly, Beschorner and Krause argue, human decisions are made with the expectation to be judged.

All three aspects are linked and distinguish human decision making from the algorithmic calculation of an optimized outcome. Thus, to pair the human moral agent with an artificial calculative agent diffuses moral agency and it unsettles the social relation necessary for the moral agent to understand themselves as such. Rubel, Castro and Pham (2019) have given this a name: agency laundering. Agency laundering is what happens 'when agents insert technological systems into their decision-making processes', which then 'obscures moral responsibility'. This succinctly captures the present challenges for human–machine teaming in warfare.

References

Amoore, L. (2020) *Cloud Ethics: Algorithms and the Attributes of Others and Ourselves*, Durham: Duke University Press.

Anders, G. (2010) *Die Antiquiertheit des Menschen – Über die Seele im Zeitalter der zweiten industriellen Revolution*, Munich: C.H. Beck.

Arendt, H. (1970) *On Violence*, San Diego: Harcourt Brace & Company.

Arendt, H. (2003) *Responsibility and Judgement*, J. Kohn (ed), New York: Schocken.

Arkin, R.C. (2009) *Governing Lethal Behavior in Autonomous Robots*, Boca Raton: CRC Press.

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F. and Rahwan, I. (2018) 'The Moral Machine experiment', *Nature*, 563(7729): 59–64. https://doi.org/10.1038/s41586-018-0637-6

Bächle, T.C. and Bareis, J. (2022). '"Autonomous weapons" as a geopolitical signifier in a national power play: analysing AI imaginaries in Chinese and US military policies', *European Journal of Futures Research*, 10(20). https://doi.org/10.1186/s40309-022-00202-w

BAE Systems (2020) 'BAE Systems selected to provide autonomy capabilities for DARPA's Squad X program', *BAE Systems News*, 2 June. Available from: https://www.baesystems.com/en/article/providing-autonomy-capabilities-for-darpa-s-squad-x-program [Accessed 3 January 2024].

Baker, D. (2022) *Should We Ban Killer Robots?*, Cambridge: Polity Press.

Beschorner, T. and Krause, F. (2021) 'Algorithms, decision making, and the human outside the code', *Morals and Machines*, 2: 79–85. https://doi.org/10.5771/2747-5174-2021-2-78

Bhuta, N., Beck, S., Geiss, R. and Kress, C. (2016) *Autonomous Weapons Systems: Law, Ethics Politics*, Cambridge: Cambridge University Press.

Bryson, J. (2010) 'Robots should be slaves', in Y. Wilks (ed) *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues*, Amsterdam: John Benjamins, pp 63–74.

Business Research Company (2023) *Autonomous Military Weapons Global Market Report 2023*. Available from: https://www.reportlinker.com/p06240604/Autonomous-Military-Weapons-Global-Market-Report.html [Accessed 3 January 2024].

Cappuccio, M.L., Galliott, J.C. and Alnajjar, F.S. (2022) 'A taste of armageddon: a virtue ethics perspective on autonomous weapons and moral injury', *Journal of Military Ethics*, 21(1): 19–38. https://doi.org/10.1080/15027570.2022.2063103

Clynes, M.E. and Kline, N.S. (1960) 'Clyborgs and space', *Astronautics*, 14(9): 26–27.

Coker, C. (2013) *Warrior Geeks: How the 21st Century Technology Is Changing the Way We Fight and Think about War*, New York: Columbia University Press.

Cummings, M.L. (2004) 'Automation bias in intelligence time critical decision support systems', *Conference Proceedings of the American Institute of Aeronautics and Astronautics*, AIAA 2004-6313, 20 September. https://doi.org/10.2514/6.2004-6313

Danaher, J. (2022) 'Tragic choices and the virtue of techno-responsibility gaps', *Philosophy & Technology*, 18(4): 299–309. https://doi.org/10.1007/s13347-022-00519-1

DARPA (2019) 'With Squad X, dismounted units partner with AI to dominate battlespace', 12 July. Available from: https://www.darpa.mil/news-events/2019-07-12 [Accessed 4 December 2024].

Deigh, J. (1996). *The Sources of Moral Agency: Essays in Moral Psychology and Freudian Theory*, Cambridge: Cambridge University Press.

Edmonds, D. (2015) *Would You Kill the Fat Man? The Trolley Problem and What Your Answer Tell Us about Right and Wrong*, Princeton: Princeton University Press.

Ekelhof, M. (2019) 'Moving beyond semantics on autonomous weapons: meaningful human control in operation', *Global Policy*, 10(3): 343–348. https://doi.org/10.1111/1758-5899.12665

Ess, C.M. (2020) 'Trust and information and communication technologies', in J. Simon (ed) *The Routledge Handbook of Trust and Philosophy*, New York: Routledge, pp 405–420.

Fischer, J.M., Kane, R., Pereboom, D. and Vargas, M. (eds) (2007) *Four Views on Free Will*, Oxford: Blackwell.

Fischer, J.M. and Ravizza, M. (1998). *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge: Cambridge University Press.

Foot, P. (1967) 'The problem of abortion and the doctrine of double effect', *Oxford Review*, 5: 1–5.

Galliott, J. (2016) *Military Robots: Mapping the Moral Landscape*, Abingdon: Routledge.

Galliott, J. (2020) 'No hands or many hands? Deproblematizing the case for lethal autonomous weapons systems', in S.C. Roach and A.E. Eckert (eds) *Moral Responsibility in Twenty-First-Century Warfare: Just War Theory and the Ethical Challenges of Autonomous Weapons Systems*, New York: State University of New York Press, pp 155–180.

Gault, M. (2023) 'Palantir demos AI to fight wars but says it will be totally ethical don't worry about it', *Vice*, 23 April. Available from: https://www.vice.com/en/article/qjvb4x/palantir-demos-ai-to-fight-wars-but-says-it-will-be-totally-ethical-dont-worry-about-it [Accessed 3 January 2024].

Gellers, J.C. (2020) *Rights for Robots: Artificial Intelligence, Animan and Environmental Law*, Abingdon: Routledge.

Government of the Netherlands (2023) 'REAIM 2023 call to action', 16 February. Available from: https://www.government.nl/documents/publications/2023/02/16/reaim-2023-call-to-action [Accessed 3 January 2024].

Gunkel, D. (2017) *The Machine Question: Critical Perspectives on AI, Robots and Ethics*, Cambridge, MA: MIT Press.

Heims, S.J. and Wiener, N. (1989) 'Introduction', in N. Wiener (ed) *The Human Use of Human Beings: Cybernetics and Society*. London: Free Association Books.

Heller, K.J. (2023). The concept of "the human" in the critique of autonomous weapons. *Harvard National Security Journal*, 15.

Himma, K.E. (2009) 'Artificial agency, consciousness and the criteria for moral agency: what properties must an artificial agent have to be a moral agent?', *Ethics and Information Technology*, 11(1): 19–20. https://doi.org/10.1007/s10676-008-9167-5

Hoijtink, M. (2022) ''Prototype warfare': innovation, optimisation and the experimental way of warfare', *European Journal of International Security*, 7(3): 322–336. https://doi.org/10.1017/eis.2022.12

Holland, M.A. (2023) *Recalibrating Assumptions on AI: Towards an Evidence-Based and Inclusive AI Policy Discourse*, London: Royal Institute of International Affairs.

Hughes, Z. (2020) 'Fog, friction, and thinking machines', *War on the Rocks*, 11 March. Available from: https://warontherocks.com/2020/03/fog-friction-and-thinking-machines/ [Accessed 3 January 2024].

International Committee of the Red Cross (2016) 'Views of the International Committee of the Red Cross (ICRC) on autonomous weapon system'. Available from: https://www.icrc.org/en/document/views-icrc-autonomous-weapon-system [Accessed 8 February 2024].

Jasanoff, S. (2015) 'Future imperfect: science, technology and the imaginations of modernity', in S. Jasanoff and S.-H. Kim (eds) *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*, Chicago: Chicago University Press, pp 1–33.

Jonas, H. (2001) *The Phenomenon of Life: Toward a Philosophical Biology*, Evanston: Northwestern University Press.

Kane, R. (2002) *The Oxford Handbook of Free Will*, Oxford: Oxford University Press.

Keenan, T. (1997) *Fables of Responsibility: Aberrations and Predicaments in Ethics and Politics*, Stanford: Stanford University Press.

Kelman, H.C. (1973) 'Violence without moral restraint: reflections on the dehumanization of victims and victimizers', *Journal of Social Issues*, 29(4): 25–61. https://doi.org/10.1111/j.1540-4560.1973.tb00102.x

Knight, W. (2023) 'Eric Schmidt is building the perfect AI war-fighting machine', *Wired*, 13 February. Available from: https://www.wired.com/story/eric-schmidt-is-building-the-perfect-ai-war-fighting-machine/ [Accessed 3 January 2024].

Lloyd, B. and Rozman, J. (2022) 'Achieving decision dominance through convergence: the US Army and JADC2', *Association of the United States Army*, February. Available from: https://www.ausa.org/publications/achieving-decision-dominance-through-convergence-us-army-and-jadc2# [Accessed 3 January 2024].

MacIntosh, D. (2021) 'Fire and forget: a moral defense of the use of autonomous weapons systems in war and peace', in J. Galliott, D. MacIntosh and J.D. Ohlin (eds) *Lethal Autonomous Weapons: Re-examining the Law and Ethics of Robotic Warfare*, Oxford University Press, pp 9–23.

MacIntyre, A. (1999) 'Social structures and their threats to moral agency', *Philosophy*, 74(3): 311–29. https://doi.org/10.1017/S0031819199000431

Markets and Markets (2023) 'Artificial intelligence (AI) in military markets (AS6116)'. Available from: https://www.marketsandmarkets.com/Market-Reports/artificial-intelligence-military-market-41793495.html [Accessed 3 January 2024].

Matthias, A. (2004) 'The responsibility gap: ascribing responsibility for the actions of learning automata', *Ethics and Information Technology*, 6: 175–83. https://doi.org/10.1007/s10676-004-3422-1

McFarland, T. (2022) 'Minimum levels of human intervention in autonomous attacks', *Journal of Conflict and Security Law*, 27(3): 387–409. https://doi.org/10.1093/jcsl/krac021

McMahan, J. (2013) 'Where to now for just war theory? Ethics, law and armed conflict podcast', *University of Oxford, Apple Podcasts*, 8 February. Available from: https://podcasts.apple.com/gb/podcast/where-to-now-for-just-war-theory/id1107584157?i=1000367485627 [Accessed 7 February 2024].

McQuillan, D. (2022) *Resisting AI: An Anti-fascist Approach to Artificial Intelligence*, Bristol: Bristol University Press.

Munch, L., Mainz, J. and Bjerring, J.-C. (2023) 'The value of responsibility gaps in algorithmic decision-making', *Ethics and Information Technology*, 25(21): 20–31. https://doi.org/10.1007/s10676-023-09699-6

Nandy, A. (1997) 'Modern science and authoritarianism: from objectivity to objectification', *Bulletin of Science Technology & Society*, 17(8): 8–12. https://doi.org/10.1177/027046769701700103

Nickel, P.I., Franssen, M. and Kroes, P. (2010) 'Can we make sense of the notion of trustworthy technology?', *Knowledge, Technology and Policy*, 23(3): 429–44. https://doi.org/10.1007/s12130-010-9124-6

O'Driscoll, C. (2022) 'Heartfelt truths: toward an existentialist ethics of war', *Review of International Studies*, 49(5): 872–884. https://doi.org/10.1017/S026021052200047X

Oimann, A.-K. (2023) 'The responsibility gap and LAWS: a critical mapping of the debate', *Philosophy & Technology*, 36(3): 1–22. https://doi.org/10.1007/s13347-022-00602-7

Palantir (Director) (2023) 'Introducing: artificial intelligence platform for defense', promotional video, April. Available from: https://www.palantir.com/platforms/aip/ [Accessed 3 January 2024].

Riesen, E. (2022) 'The moral case for the development and use of autonomous weapon systems', *Journal of Military Ethics*, 21(2): 132–150. https://doi.org/10.1080/15027570.2022.2124022

Roberson, T., Bornstein, S., Liivoja, R., Ng, S., Scholz, J. and Devitt, K. (2022) 'A method for ethical AI in defence: a case study of developing trustworthy autonomous systems', *Journal of Responsible Technology*, 11(100036). https://doi.org/10.1016/j.jrt.2022.100036

Rubel, A., Castro, C. and Pham, A. (2019) 'Agency laundering and algorithmic systems', in N. Taylor, C. Christian-Lamb, M. Martin and B. Nardi (eds), *Information in Contemporary Society*, Berlin: Springer Nature, pp 590–598.

Schmidt, E., Work, B., Catz, S., Chien, S., Darby, C., Ford, K. and Griffiths, J.-M. (2021) 'National Security Commission on Artificial Intelligence final report', *National Security Commission on Artificial Intelligence (NSCAI)*. Available from: nscai_full_report_digital.04d6b124173c.pdf (foleon.com) [Accessed 7 February 2024].

Schwarz, E. (2017) 'Hybridity and humility: what of the human in posthuman security?', in C. Eroukhmanoff and M. Harker (eds), *Reflections on the Posthuman in International Relations: The Anthropocene, Security and Ecology*, Bristol: E-International Relations Publishing, pp 29–38.

Schwarz, E. (2018) *Death Machines: The Ethics of Violent Technologies*, Manchester: Manchester University Press.

Schwarz, E. (2019) 'Günther Anders in Silicon Valley: artificial intelligence and moral atrophy', *Thesis Eleven*, 153(1): 94–112. https://doi.org/10.1177/0725513619863854

Schwarz, E. (2021) 'Autonomous weapons systems, artificial intelligence and the problem of meaningful human control', *Philosophical Journal of Conflict and Violence*, V(1): 53–72. https://.doi.org/10.22618/tp.pjcv.20215.1.139004

Schwarz, E. (2024) 'Trolleyology: algorithmic ethics for killer robots', in D. Gunkel (ed) *Handbook on the Ethics of Artificial Intelligence*, Cheltenham: Edward Elgar, pp 83–96.

Shepardson, D. (2019) 'In review of fatal Arizona crash, US agency says Uber software had flaws', Reuters, 6 November. Available from: https://www.reuters.com/article/us-uber-crash-idUSKBN1XF2HA [Accessed 3 January 2024].

Simpson, T.W. and Müller, V.C. (2015) 'Just war and robots killing', *Philosophical Quarterly*, 66(263): 302–322. https://doi.org/10.1093/pq/pqv075

Smiley, L. (2023) 'The legal saga of Uber's fatal self-driving car crash is over', *Wired*, 28 July. Available from: https://www.wired.com/story/ubers-fatal-self-driving-car-crash-saga-over-operator-avoids-prison/ [Accessed 3 January 2024].

Sparrow, R. (2007) 'Killer robots', *Journal of Applied Philosophy*, 24(1): 62–77. https://doi.org/10.1111/j.1468-5930.2007.00346.x

Suchman, L. (2023) 'The uncontroversial "thingness" of AI', *Big Data & Society*, 10(2): 1–5. https://doi.org/10.1177/20539517231206794

Sullins, J.P. (2006) 'When is a robot a moral agent?', *International Review of Information Ethics*, 6(12): 24–29.

Trouver, S.C. (2012) 'Taking mines seriously: mine warfare in China's near seas', *Naval War College Review*, 65(2): 30–66.

Turing, A. (1950) 'Computing machinery and intelligence', *Mind*, LIX(236): 433–460).

UK Ministry of Defense (2022) 'Ambitious, safe, responsible: our approach to the delivery of AI-enabled capability in defence', policy paper. Available from: https://www.gov.uk/government/publications/ambitious-safe-responsible-our-approach-to-the-delivery-of-ai-enabled-capability-in-defence/ambitious-safe-responsible-our-approach-to-the-delivery-of-ai-enabled-capability-in-defence [Accessed 3 January 2024].

Umbrello, S. (2022) *Designed for Death: Controlling Killer Robots*, Budapest: Trivent Publishing.

US Department of Defense (2022) 'Summary of the Join All-Domain Command & Control (JADC2) strategy', 17 March. Available from:. https://media.defense.gov/2022/Mar/17/2002958406/-1/-1/1/SUMMARY-OF-THE-JOINT-ALL-DOMAIN-COMMAND-AND-CONTROL-STRATEGY.PDF [Accessed 3 January 2024].

US Department of Defense (2020) 'DoD adopts ethical principles for artificial intelligence', 24 February. Available from: https://www.defense.gov/News/Releases/release/article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/ [Accessed 3 January 2024].

Verdiesen, I. (2019) 'The design of human oversight in autonomous weapon systems', in V. Conitzer, S. Kambhampati, S. Koenig, F. Rossi and B. Schnabel (eds) *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics and Society*, Association for Computing Machinery (ACM), pp 388–389. https://doi.org/10.1145/3278721.3278785[6]

Walker, P. (2021) 'Leadership challenges from the deployment of lethal autonomous weapon systems', *RUSI Journal*, 166(1): 10–21. https://doi.org/10.1080/03071847.2021.1915702

Walsh, T. (2022) 'The problem with artificial (general) intelligence in warfare', *Center for International Governance Innovation (CIGI)*, 28 November, https://www.cigionline.org/articles/the-problem-with-artificial-general-intelligence-in-warfare/ [Accessed 3 January 2024].

Wiener, N. (1954) *The Human Use of Human Beings: Cybernetics and Society*, Boston: Da Capo Press.

Legitimizing and Contesting Lethal Autonomous Weapons Systems in Japan: A Multilayered Analysis of Public Discourse

Bernhard Seidl

Introduction

According to the French philosopher Michel Foucault, discourses are social processes that systematically form the very subjects they talk about (Foucault, 2004: 54), and true to this assertion, the most basic assumption of discourse theory concerns the relation between language, knowledge and action: through language, knowledge is stored and shared. Knowledge in turn is the prerequisite for action, and action is what creates social reality. Discourses are therefore not merely passive mirrors of a society, but rather creative processes that motivate and guide social action (Jäger, 2001: 34–38; Keller, 2013: 23). Consequently, analysing a discourse gives insight not only into the intersubjective social realities of a topic, but also into the various factors involved in the process of its realization.

The discourse on lethal autonomous weapon systems (LAWS) is no exception, and this chapter aims to shed some light on how the social reality regarding LAWS is shaped by sociocultural knowledge, norms and imaginaries. Discourse often develops at the intersection of several key issues, and for the wider discourse on LAWS, two such issues are robotics and security identity, the 'collectively held principles … regarding the appropriate role of state action in the security arena' (Oros, 2017: 18).

Japan was selected as the focus of this analysis for several reasons. First, despite many domestic challenges, it remains the world's third-largest economy as of 2023, making it a key geopolitical player in East Asia and the Indo-Pacific. Moreover, in both of the key areas identified previously, Japan can be seen as uniquely positioned: it is often seen as a 'robot nation' both domestically and internationally, and regardless of the factuality of this label, this is evident in the government's visions of a 'Society 5.0' (2016) driven by AI and robots, industry initiatives such as the 'New Robot Strategy' (2015), and popular culture. Finally, Japan presents an intriguing case as a 'major military power with without a "military"' (Iwamoto, 2021: 2): a nation that continues to define itself as fundamentally opposed to war as a means of conflict resolution, and that by definition has no army, yet has increased its military budget for 2023 to a remarkable $49 billion (Japanese Ministry of Defence, 2022).

The dominant security identity of a nation also determines how war-making capabilities such as the development and adoption of weapons are argued for or against. Consequently, the maintenance of military forces, the rules of engagement for those forces, and the use of national resources for military technology must be reconciled with it. In the case of Japan, for example, this has been evident in the government's repeated assertions of its commitment to international peace-building efforts, its emphasis on building up military power for the sole purpose of self-defence, or its focus on funding civilian-military dual-use technologies. Against this backdrop, several questions can be raised regarding the LAWS discourse in Japan: How is a new weapons technology such as LAWS argued for or against? What rationalizations and justifications, metaphors or symbolisms are used, and to what extent are they influenced by tenets of the security identity?

In order to address these questions, a discourse analysis following the structure proposed by Siegfried Jäger will be carried out: based on a representative corpus, macrolayers and microlayers of text are analysed, systematically exploring linguistic patterns and connections, and situating the findings in social, cultural and historical contexts (Jäger, 2015). Focusing on political, civil society and mass media discourse, this chapter examines public discourse – that is, discourse produced in or intended for the public sphere – aiming to provide a comprehensive review and critical discussion of publicly available resources produced by the Japanese government, political parties and nongovernmental organizations (NGOs). This is combined with a corpus-based analysis of newspaper articles, on the assumption that newspapers remain an important medium of information transfer and public opinion formation in Japan.

By situating the findings in the context of Japan's changing security identity and 'making visible the interconnectedness of things' (Fairclough, cited in Wodak, 2001: 2), this chapter aims to explore how LAWS technology, with its potential to fundamentally change the face of armed conflict, is discussed and framed in Japanese public discourse through the identification of topics and linguistic devices (van Dijk, 2001: 102), thereby providing an insight into the various realities and imaginaries that inform the discourse on LAWS in Japan.

The following section first provides an overview of Japan's postwar security policy and recent trends, and proceeds to examine the academic

LAWs discourse to offer important context for the subsequent analysis. The subsequent section will then explore the policy discourse on LAWS before the analysis is extended to the journalistic discourse and an examination will be given of the key issues, voices and positions of a discourse that is driven by questions of what is technologically feasible, ethically permissible and pragmatically achievable.

1. Historical context and academic discourse of Japan's security identity

Japan's present constitution was drafted under the supervision of US General Douglas MacArthur, who was in charge of demilitarization and democratization efforts in Japan, and came into force in 1947. It is often referred to as a 'Peace Constitution' due to its Article 9, which forbids Japan from having an army, but has been interpreted as not precluding the right to self-defence. Japan has therefore maintained 'Japan Self-Defence Forces' (JSDF) with a limited mandate since 1954 with support from the US, a military ally since the 1951 Japan-US Security Treaty: having succeeded in building a 'picture-book' democracy, the US now needed a partner in the region as part of its anti-communist containment strategy (Berkofsky, 2010: 24).

While the US-Japanese security treaty and its revision of 1960 has remained a contentious issue in Japan (Oros, 2017), the postwar period was nevertheless shaped by the 'Yoshida Doctrine': a focus on economic growth while minimizing defence spending, maintaining minimal armament and relying on the US for military security. With the largest US military presence outside of its national territory (Haddad and Hussein, 2021), Japan still is an important forward base for the US in the region, although there has also been criticism of Japan hitching a 'free ride' on the back of the US military presence (Oros, 2017: 59). This was recently brought into focus by US President Donald Trump's criticism of the security treaty as 'unfair' (Reuters, 2019), which led to an increase of Japan's financial contribution to US bases in Japan under the Biden administration (Seligman and Gramer, 2019; Matsuyama and Sato, 2021).

To better situate the LAWS discourse within the larger discourse on national security, the following section provides an overview of the development of Japan's security identity from the postwar period to the present.

1.1 Shifts in the Japanese security identity

The early postwar period was the formative time of what Andrew Oros calls the Japanese security identity of domestic antimilitarism (SIDA), which can be seen as the sociopolitical manifestation of the constitution and lingering anti-militaristic sentiment in the aftermath of a brutal war. Oros contends that the SIDA rests on three tenets he summarizes as follows:

1. Japan will possess no traditional armed forces
2. There will be no use of force by Japan except in self-defence
3. No Japanese participation in foreign wars (Oros, 2015: 140)

However, since the early 1990s, the SIDA has come increasingly under pressure from events and circumstances that have led to a growing sense of need to re-evaluate the passive postwar security policy, resulting in a shift towards a more assertive international state identity (Singh, 2013) that had to be reconciled with the persistent 'peace state' master narrative.

Examples of such events include the 'Gulf War trauma' (Maslow, 2015: 739) of the early 1990s, when the self-imposed limits of security policy prevented Japan from actively contributing to United Nations (UN) Security Council Resolution 678, resulting in diplomatic and journalistic scorn for Japan's tardy 'checkbook diplomacy' (Iwamoto, 2021: 25) or the 'Taepodong shock' (Kabata, 2018: 32) when North Korea successfully fired a ballistic missile over the Japanese islands in 1998, the first of many such missile tests that North Korea has conducted to date. Yet, in addition, factors such as the rise of China as a regional anti-Japanese 'Other' since the late 1990s (Hook et al, 2015: 20; Gustafson, 2016: 188), fuelled in part by disputes over the Senkaku/Diaoyu Islands as well as similar territorial issues with South Korea and Russia, brought Japan's lack of military power and its dependence on the US into focus. In addition, Japan's rapidly ageing society placed an ever-increasing burden on the social security budget, which by 2015 had become a major source of fiscal imbalance and also resulted in a rapidly shrinking recruitment pool for the JSDF (Funabashi, 2016).

In response to such 'risk narratives' (cf. Hook et al, 2015), Japan gradually adapted its international security policy, starting with a bill to allow the JSDF to participate in international peace-keeping operations in 1992, followed by strengthening military partnerships and updating military equipment (Maslow, 2015: 740). These changes came increasingly into international focus during Abe Shinzō's tenure as Prime Minister from 2012 to 2020 under the slogan of 'taking back Japan' (*nippon o torimodosu*). Among many structural changes, notably bills that concentrated the fragmented structures of foreign policy decision making in the Prime Minister's Office, Abe hoped to eliminate Article 9, which the nationalist LDP[1] elites he represented had long seen as humiliating (Hughes et al, 2021: 129–130).

However, these plans not only met with fierce opposition in the Diet (Oros, 2017: 106), but also proved unpopular with the voters: Abe misjudged

[1] Liberal Democratic Party; Jiyū minshutō.

how deeply the distrust for anything remotely associated with militarization was still rooted in civil society, and especially among the elderly (Funabashi, 2016: 28). Instead, the LDP settled for a reinterpretation of Article 9 via the *Collective Defense Bill*, which is seen by the government as covered by the Right to Self-Defence as stipulated in Chapter VII, Article 51 of the United Nations Charter, and while this too was widely and controversially debated in Japan (Oros, 2017: 3), the LDP managed to push it through with the support of its coalition partner NKP,[2] which is opposed to a revision of the constitution (Shio, 2016).

The reinterpretation of Article 9 led to a heated debate in Japan and was also viewed with suspicion by neighbouring countries as the first signs of a remilitarization, with the sometimes blatant historical relativism and revisionism of Japanese leaders contributing to a dubious overall image in this period of gradual identity shift. While Abe's policies have sought to enhance Japan's military capabilities through modernization and closer international relations, Japan has not been able to escape the discursive constraints of the SIDA. This is evident in the mantra-like assertions of pacifism and commitment to peace in official documents, but while there is a move away from inward-looking, unilateral pacifism towards a more assertive security policy through what is rather ambiguously termed 'proactive pacifism' (*sekkyokuteki heiwashugi*) in the National Defence Program Guidelines for 2019 and Beyond. While this new doctrine has received much criticism in academic circles, such as from legal scholar Okuno Tsunehisa, who describes it as a security policy that essentially 'seeks to promote peace through military strength and support for US wars' (Okuno, 2023: 140), it is also worth noting that the SIDA was never an explicitly pacifist identity to begin with (Oros, 2017: 56).

The Russian invasion of Ukraine in 2022 has undoubtedly acted as a catalyst for an increasingly assertive security policy, including an ambitious plan to increase the defence budget, although the sustainability of such a budget seems doubtful (Bosack, 2023; Reuters, 2023). Furthermore, information warfare and transnational cyberattacks are 'blurring the line between contingency and peacetime', as the 2022 National Security Strategy contends (Cabinet Secretariat of Japan, 2022: 2), implicitly further questioning where Japan's right to self-defence begins. It is therefore worth examining how this continued identity shift has manifested in the LAWS debate as an example of military-focused discourse. In order to provide further background for the analysis that follows, and to highlight the influence of security identity on the wider discourse, the following section gives an overview of the academic debate on LAWS in Japan.

[2] New Komeito Party; Kōmeitō can be loosely translated as 'party of fair and honest politics'.

1.2 Academic discourse on LAWS in Japan

The scholarly discussion of LAWS in Japan started soon after the Heyns Report (Heyns, 2013) was published, and the amount of research on LAWS-related topics in Japanese humanities and its fringe areas like robot or AI ethics points to a lively academic debate. One interesting aspect is that several of its key figures[3] have also been active as intermediaries between the academic world and civil society, for example, as lecturers at nonprofit organization events or as members of nongovernmental think tanks and research societies.[4] Furthermore, some have also been called upon by the government as experts, such as Fukui Yasuhito and Yoshida Yasuyuki, who were appointed to the Japanese GGE[5] delegation in 2019, or Satō Heigo, who was invited to the Committee on Foreign Affairs and Defence in the Japanese Parliament as an external expert on LAWS (NDL, 2019). It can therefore be assumed that the scholarly discussion of LAWS in Japan, which is mainly focuses on laws, ethics and political studies, is noticed by policy makers and contributes to shaping the public discourse on LAWS well beyond the proverbial walls of academic seclusion. The question of the ethicality of LAWS has been central from the beginning (Kukita et al, 2014; Kanzaki, 2021) and has remained a main topic since, with the question of culpability of robots and AI sometimes extended to the Japanese legal framework (Shinpo, 2017, 2020). The second central topic of research is international humanitarian law (IHL) and its applicability to LAWS (for example, Satō, 2014; Fukui, 2017; Yoshida, 2020; Iwamoto, 2022). Here, we also find the bulk of the literature commenting on the efforts for LAWS regulation at the CCW/GGE. Other topics include the stance of the Japanese government (for example, Ueno, 2019), the role of NGOs (Satō, 2020) or the question of military-civilian dual use of technology and the ethicality of doing research for military purposes (Kabata, 2018; Takahashi, 2021).

The latter is interesting also as a manifestation of the SIDA in academia, and an example of a clash of interpretations between the government and the academic community, which has a long tradition of distancing itself from any military research. A case in point is the declarations published by the Science Council of Japan[6] (SCJ), founded in 1949 by order of the

[3] Notably Fukui Yasuhito, Satō Heigo and Iwamoto Seigo.

[4] Such as the Center for Information on Security Trade Control (CISTEC), the Arms and Civil Society Research Forum 'Militarization and Civil Society' and the Japanese Association of Disarmament Studies (JADS).

[5] Group of Governmental Experts on Lethal Autonomous Weapons Systems meeting under the umbrella of the UN Convention on Certain Conventional Weapons

[6] Nihon gakujutsu kaigi.

government, as 'the representative organization of the Japanese scientific community ranging over all fields of sciences' (SCJ, 2022).

In this capacity, the SCJ published two declarations outlining the stance of the Japanese science community regarding research for military purposes in the postwar period: the 1950 'Declaration of intent to never pursue research that is to be used for war' and the 1954 'Declaration to not pursue research that is to be used for military purposes' (Kabata, 2018: 32). These declarations, although without any legal power, have had a lasting and defining effect on Japanese military technology research, which stayed largely limited to private or state-owned companies.

With the gradual abandonment of the postwar security policy and a sense of urgency to modernize the JSDF (Hughes, 2011: 476), the Ministry of Defense (MOD) started to look for ways of harnessing the potential of Japanese cutting-edge research for military applications. To this end, the dual-use of technology paradigm has become an important instrument for the MOD to circumvent budgetary and ideological limitations: notwithstanding military applicability, the research is not considered 'military' and therefore does not count as part of the MOD's budget (Kallender and Hughes, 2019: 191).

However, as divisions between civilian and (potentially) military technology, especially in the sector of robotics and AI, grow less distinct, it becomes increasingly difficult to judge the ultimate ethicality of one's research. The SCJ's 2017 'Statement on military security research' may be seen to reflect this; while criticized by some as too meek of wording in comparison to its 1950/1954 predecessors (Kabata, 2018: 37), it nevertheless urges research facilities to publicly account for the possibility of their research ending up being used for military applications.

The simmering dispute between the SCJ and the government has manifested not only in publications like astrophysicist Ikeuchi Satoru's 2019 book *Why Scientists Must Not Engage in Military Research*,[7] but also in the lawful yet unprecedented and widely criticized rejection of several SCJ members by Abe's successor Suga on vague grounds that are believed to also be connected to the SCJ's critical stance on military technology research (Takahashi, 2021: 71). In December 2022, the government announced plans to strengthen defence capabilities with a focus on advanced technologies; while the MOD aims to achieve this by strengthening cooperation between public and private researchers and policymakers (Japanese Ministry of Defence, 2022: 27), it remains to be seen how enthusiastically Japan's scientific community will participate in these plans.

7 *Kagakusha wa naze gunjikenkyū ni te o somete wa ikenai ka.*

2. Lethal autonomous weapon systems in Japanese policy-making discourse

While the academic debate on LAWS mainly focuses on the ethical and legal implications, the political discourse is situated between the government's desire to harness the potential of LAWS technology, the need to frame this desire in a way that is compatible with the Japanese peace-oriented security identity, and NGOs trying to influence the policy discourse through interacting with the public as well as political parties. To achieve a comprehensive overview, the material for the following subsections was gathered primarily from NGO, government, and political party[8] websites, and the National Diet Proceedings Search System of the National Diet Library. Specific websites were identified through iterative internet searches using synonyms for LAWS,[9] and further scrutinized by using domain- or subdomain-specific searches. The results were then collected, manually tagged and interpreted.

2.1 Political parties

With brief interludes, the national conservative LDP has been the defining political force in post-Second World War Japan. Since December 2012, the Japanese government has been led by a coalition of the LDP and the much less influential NKP under three prime ministers. Akimoto (2019: 316) notes that LAWS were the subject of debates in the Japanese Diet ten times between 2015 and 2019, with the main subjects being ethical concerns and calls to regulate or ban LAWS. While there was a similar trend with a comparable number (11) of discussions between the second half of 2019 and June 2023, the focus lies on urging the government to take a leadership role in international regulatory efforts and discussing issues such as accelerating AI development in civilian and defence contexts, as well as defining the limitations of autonomous AI.

Overall, the government's reasoning, repeated by officials in most discussions, is not challenged: Japan must strengthen national security; Japan needs automation to deal with the problem of population decline; the line between military and civilian technology is becoming less distinct; Japan needs to focus on technologies such as AI and automation. However, there are several debates in which law makers challenge details of the government's arguments, such as whether Japan can actually afford to forgo LAWS when

8 Limited to those present in the National Diet as of 2023.

9 LAWS, *jiritsugata/jiritsusei chishi heiki shisutemu, AI heiki, jinkō chinō heiki, killer robot, kirā robotto, satsujin robotto*.

nations such as China may have no such qualms, or whether the government's pledge extends to AI systems designed to repel cyberattacks (NDL, 2021).

While the Japanese government transports its stance on LAWS to the public through its web portals (see the next section), most parties were not found to be as publicly communicative on the issue. Neither the LDP nor the main opposition party CDP[10] has anything to say about LAWS on their websites or in their election manifestos. With the exception of the JCP,[11] which discusses LAWS in a critical tone in three articles of its party newspaper, and especially the NKP, the other parties were found to be similarly reticent about LAWS. This makes the extent of the NKP's engagement with the LAWS issue all the more striking. The NKP is often called the political arm of the Sōka Gakkai, a Buddhist society with over 8 million members in Japan (DWD, 2021), and incidentally a member of the multi-NGO initiative Campaign to Stop Killer Robots (CSKR). As a party rooted in Buddhist thought, the NKP has a strong pacifist orientation and takes a firm stance on disarmament issues. The self-styled 'Party of Peace' proclaims to 'guide the ruling party with down-to-earth pacifism' (NKP, n.d.); this down-to-earth side of the NKP's pacifism being evident in its support for the 1992 PKO Cooperation Law and, more recently, the much-discussed 2014 revision of the JSDF's operational guidelines (*shin san'yōken*).

The NKP began to publicly address the issue of LAWS in 2018, when party leader Yamaguchi Naoto met with Kenneth Roth from Human Rights Watch (HRW). The report on the meeting quotes Yamaguchi as saying: 'It is necessary to establish regulations to prevent the development and export of killer robots. We will take a humanitarian stand [on this issue]' (NKP, 2018). LAWS are explicitly discussed in the two following election manifestos (2019 and 2022), both of which describe the party's position on LAWS in terms largely identical to the official government's stance, and between 2018 and 2022, numerous articles containing related information were published on the NKP web portal, including reports on GGE talks, the participation of NKP representatives in NGO-organized discussion events (AAR Japan, 2018; TVAC, 2018), expert interviews and several updates on a LAWS-related NKP working group.[12] On several of these webpages, the NKP not so subtly claims issue ownership as the first political party in Japan to deal with the topic by adding the footer 'Ahead of other parties: Studying LAWS, advising the Government' (for example, NKP, 2019).

[10] Constitutional Democratic Party of Japan (Rikken minshutō).

[11] Japanese Communist Party (Nihon Kyōsantō).

[12] Project team for investigating LAWS development regulations (LAWS kaihatsu seigen ni kan suru kentō purojekuto chīmu); since there have been no news on the PT since 2019, it can be assumed to be defunct.

This and similar statements serve to highlight the NKP's active role in LAWS-related policy making, while setting the party apart from the government of which it actually is a part. While the terminology used for LAWS by the NKP ranges from technical terms to 'killer robot', no clear trend could be identified. However, in terms of content, there is a clear focus on ethical issues and a call for regulations.

2.2 The Japanese government

The Japanese government's stance on LAWS regulation falls into what Iwamoto Seigo calls the 'non-legal instruments camp', meaning that Japan does not call for explicit disarmament or IHL conventions, but rather for legally nonbinding instruments like codes of conduct or other 'soft laws', such as the Guiding Principles for the Development and Use of LAWS proposed by the CCW-GGE[13] in 2019 (Iwamoto, 2022: 11).

This position has hardly changed since Sano Toshio's speech at the first CCW meeting on LAWS in 2014, which remains one of the most detailed statements. The key points of this speech are that Japan 'may continue researching and developing non-lethal autonomous technology for defence purposes', but 'is not convinced of the need to develop "fully" [*sic*] lethal autonomous weapon systems which is [*sic*] completely out of control of human intervention' (Sano, 2014). LAWS are defined in the address as 'fully lethal autonomous weapon systems that, once activated, can effectively select and engage a target without human intervention', which are considered incompatible with IHL; however, the 'fully' tends to be dropped in less declarative texts and in discussions.

The 2019 *Working Paper* submitted by Japan at the GGE states that it 'is indispensable that any lethal weapon system … be accompanied with meaningful human control (for example, by ensuring proper operation by a person sufficiently informed of the weapons systems in use)' (Japanese Ministry of Foreign Affairs, 2019a: 3), but ultimately, the government's understanding of existing systems as having 'a significant degree of human involvement' and therefore not being subject to regulation can be seen as reserving the option of using such systems in the future (Ueno 2019: 156), a path that has already been initiated as of 2023 (see the section on current and future involvement of autonomous AIs in warfare and national security later on in the chapter). This is accompanied by the position that, given the demographic challenges, 'Japan's defence capability cannot be secured without robust pragmatism[14] that is not bound by the past. Given this, autonomous

[13] Group of Governmental Experts on Lethal Autonomous Weapons Systems meeting under the umbrella of the UN Convention on Certain Conventional Weapons

[14] *Tettei shita gōrika.*

weapon systems that are sensible in terms of labour and manpower saving will be necessary to protect Japan's peace' (Ueno, 2019: 158).

Another key theme in the Japanese stance is to emphasize the importance of recognizing the benefits of autonomous technology in civilian applications such as healthcare, welfare and rescue. In Sano's 2014 address, this is illustrated by the use of robots for search and rescue operations and the removal of radioactive debris in the aftermath of the Fukushima disaster in 2011.

The government communicates its stance to the public through three major channels: the Ministry of Foreign Affairs (MOFA), the Ministry of Defense (MOD) and the Delegation of Japan to the Conference on Disarmament. The latter provides the least amount of information, with only one information page on the CCW stating the Japanese government's stance on LAWS, and two short reports.

On the MOFA web portal, there are also some LAWS-related news articles, but the bulk of the LAWS material available to the public (15 pages) provides a detailed account of Japan's position, and a transparent documentation of regulatory efforts from 2014 onwards, including the role of NGOs, while highlighting Japanese contributions; however, these pages are five navigational layers away from the main web portal, making it hard to access the information without knowing exactly where to look (MOFA, 2023).

In comparison, the resources on the MOD's web portal are limited to a handful of brief mentions of LAWS *inter alia* on news pages. However, the MOD also publishes a digital version of the *Whitebook of Defence*,[15] a highly visible text geared towards public accounting. LAWS are included since the 2014 edition, in which a sizable column discusses unmanned vehicles (UVs); LAWS are explained as a future but likely development. Interestingly, although humanoid robots are only briefly mentioned as a possible development, the only image in the column features Boston Dynamic's humanoid Atlas robot with its arms raised in a vaguely threatening pose (MOD, 2014).

All editions since 2014 have mentioned regulatory efforts, briefly in the context of New Military Technologies and/or in slightly more detail under Arms Control and Anti-Proliferation Initiatives, with the 2019 edition being the first to include Japan's official stance on LAWS, with a qualifying statement: 'However, weapon systems possessing autonomy may be beneficial from a security perspective taking into consideration issues such as the reduction of human error and the saving of manpower' (MOD, 2019). The 2020 edition describes AI-powered UVs as a 'game changer' technology that is 'expected to change the face of armed conflict' (MOD, 2020); however,

[15] The official English title is *Defense of Japan (Annual White Paper).*

the actual term 'LAWS' is not used in this context, and similar passages are absent from the 2021 and 2022 editions.

2.3 Nongovernmental organizations campaigning for lethal autonomous weapon systems regulations and bans

The CSKR, an international platform composed of more than 180 international NGOs (CSKR 2022), including five Japanese groups,[16] is the spearhead of anti-LAWS NGO activity in Japan. It was founded in 2013 by 11 NGOs, including the Japanese organization Association for Aid and Relief (AAR),[17] which is also a member of its steering committee. In addition to local NGOs, international ones like the International Committee of the Red Cross (ICRC) and HRW are also active, with the former in particular providing extensive Japanese material online (cf. ICRC, 2022).

The earliest example of public NGO activity regarding LAWS is a lecture entitled 'Towards a ban of killer robots', organized in October 2013 by AAR/CSKR, and information on similar events comprise most of what Japanese NGO websites offer in terms of LAWS-related information. In a 2020 talk, a CSKR Japan representative lists 13 such public lectures, study groups and symposia on LAWS until 2019. The same talk summarizes the CSKR's expectations for the Japanese government: 'Addressing the main issues of LAWS at an international conference in Japan, especially deepening the discussion on human control; civil society participation in official talks; Japanese leadership' (Sakurai, 2020: 9).

Other examples of NGO activities include an address by AAR member Osa Yukie at the 2016 GGE, public film screenings (Sōka gakkai, 2022) or the production of printed information material. A notable example of the latter is an AAR booklet entitled *We Don't Need 'Killer Robots' in This World!*,[18] intended as teaching material in schools; the AAR also offers to send a representative to schools to discuss the topic based on the booklet (AAR Japan, 2016). A similar project, a manga book entitled *Manga Introduction: Are the Killer Robots Coming?!*,[19] was launched in 2018 by Peace Boat, a rather unique Japanese NGO that promotes peace and human rights through a passenger ship travelling around the world. The group describes the motivation for the manga as follows:

> Peace Boat has been actively involved in campaigns to ban inhumane weapons … In recent years, NGOs calling for a ban or regulation

[16] Association for Aid and Relief (AAR), Human Rights Watch Japan (HRW), Human Rights Now, International Student Conference, and Soka Gakkai International.

[17] The Japanese name of the group is Nanmin o tasukeru kai (Group for Helping Refugees).

[18] *Kono sekai ni 'satsujin robotto' wa iranai!*

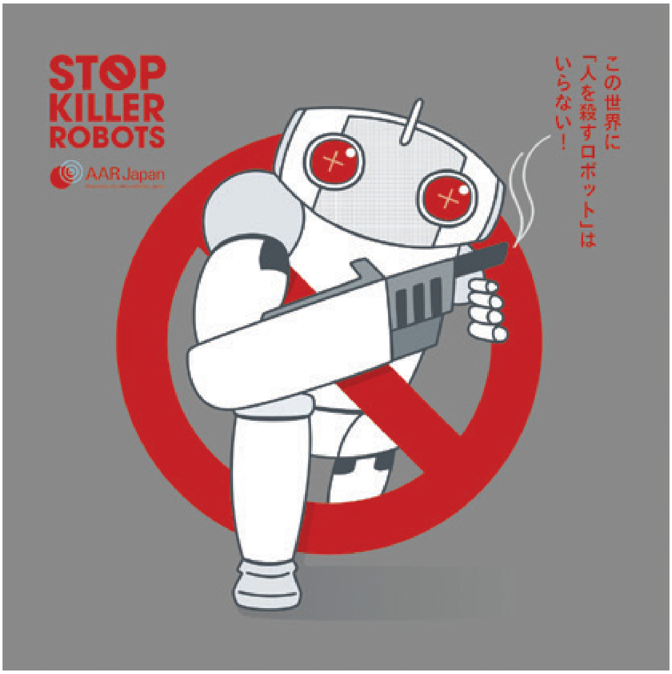
[19] *Manga nyūmon satsujin robotto ga yattekuru!?*

> of ‘killer robots’ and ‘armed drones’ have become more active internationally. However, the issue has remained largely unknown in Japan. This is why we have published this book in the form of a manga-based introductory book. (PB, 2018)

These publications also allow a glimpse into the visual side of the discourse: While the AAR booklet’s cover (see Figure 1) depicts a stylized humanoid robot with red eyes and a smoking gun moving towards the reader, the cover of Peace Boat’s manga (see Figure 2) features disjointed renditions of existing machines and two of the protagonists. Despite the absence of anthropomorphic robots or potentially sensationalist imagery, the prominent use of the word ‘killer’ (殺人 *satsujin*) in the largest font on the cover and the ‘!?’ added to the title still evoke a sense of menace.

Satō Hitoshi notes that in comparison to anti-landmine campaigns in the past, NGOs campaigning for a ban of LAWS development face several difficulties. Not only is there a lack of influential nations to lead a campaign for a ban, but impressing the need for action regarding a thing of the future with no tangible effect on the here and now on the general populace is also difficult. Furthermore, the abstract concept of an ‘AI threat’

Figure 1: We don’t need ‘killer robots’ in this world!



Source: AAR Japan (2014). Used with kind permission of AAR Japan.

Figure 2: Manga introduction: are the killer robots coming?!

Source: Cover illustration by Niina Akihiko, taken from Hatakeyama and Kawasaki (2018)

is difficult to communicate, which is why NGOs need to rely on 'imagined scenarios and stories' (Satō, 2020: 46), visible in the tendency to focus on anthropomorphic robots and 'what if' scenarios, which may explain the frequent use of the term 'killer robot' in NGO texts. For example, the home page of the 2013 event mentioned previously refers to LAWS exclusively as killer robots, asking 'Is a future coming in which 'killer robots' walk around and kill people?' (AAR, 2013), but the same tendency is visible in more recent examples too.

As NGO campaigns are geared towards banning or strictly regulating LAWS, it may be desirable to evoke a threatening image, which is arguably much harder to convey through technical-sounding language. This may also explain why NGOs frequently project an anthropomorphic image of LAWS through language (as in killer robots 'walking around') or visual representations. However, with increasing media coverage of the military use of unmanned vehicles, the debate appears to be taking on a more tangible dimension (Iwamoto, 2022: 5); in particular, the ongoing war in Ukraine, in which loitering munitions,[20] often referred to as 'kamikaze drones' in the media (Nikkei, 2022),[21] play a crucial role may help to bring the issue of autonomous weapons to the attention of a wider public.

3. The newspaper-mediated discourse on lethal autonomous weapon systems in Japan

Daily newspapers inform their readers on a wide range of topics, without the need for a specific search; while there is a lot of information to be found on LAWS in dedicated books and on websites, it stands to reason that for many people, newspapers and other nonspecific mass media will be the first and perhaps only source of information on the topic. While Japanese newspapers are struggling to compete with other sources of information, especially social media, Japan still boasts very high circulation numbers (Sawa and Saisho, 2022). Studies published by the Japanese Newspaper Association suggest that newspapers are still widely considered the most trustworthy and vital source of information, and with 60 per cent of the adult population consuming newspapers in print or digital form on a regular basis, they continue to constitute a major factor of news dissemination (Japanese Newspaper Association, 2021, 2023). Overall, as a mass medium equally accessible to all demographic groups, including the elderly, who are often less digitally literate but who represent a large demographic group in Japan's 'super-aged society' (Think Tank European Parliament, 2020), newspapers provide a good opportunity to examine the broader public discourse.

There are five Japanese national newspapers, the biggest three of which have been used in this study: *Yomiuri Shinbun*, *Asahi Shinbun* and *Mainichi Shinbun*[22] (ranked by sales; cf. Shoko-Ranking, 2020). The discourse became manifest rather abruptly in 2013, sparked by reports on the first CCW meeting on LAWS, and has continued steadily. Sampling was stopped in April 2023, so the corpus contains ten years of newspaper coverage on LAWS. With large discourses, it is typically necessary to selectively sample the data

[20] Potentially semi-autonomous one-way drones; the use of such weapons in Libya was judged by the UN to be the first documented use of LAWS in the field (UN, 2021: 17).

[21] Incidentally, the term is also used by the Japanese media in an example of linguistic re-import. It does not seem to be seen as culturally insensitive, maybe also owing to the fact that the suicidal dive bombers of the Pacific War are commonly called *tokkōtai* ('special attack units'), not *kamikaze*, in Japanese.

[22] *Shinbun* ('newspaper') will be dropped from Japanese newspapers' names in this chapter.

to produce a manageable yet still representative 'virtual corpus' (Landwehr, 2001: 107), but for the topic of LAWS, the total number of 171 relevant digital articles (*Yomiuri* 45 articles [26 per cent], *Asahi* and *Mainichi* both 63 [37 per cent]) was found to be perfectly manageable.

In order to build the newspaper corpus, key terms were extracted from the literature and the work-in-progress corpus, leading to many iterations of database scouring until no further potentially relevant results could be found. The results were checked for relevance via key term highlighting and partial reading, and articles deemed irrelevant or insubstantial were discarded.

For the analysis of the corpus, the text mining software KhCoder was used. It provides a versatile set of tools for analysing and visualizing large sets of text data, and operates on and with several units of analysis and categories of words, which are noted in brackets where this is deemed necessary.[23] Besides simple lemmatic tokens, KhCoder can also use 'codes', which represent collections of words alongside conditional operators and which will be marked with an asterisk in text (for example, *LAWS).

The main KhCoder operations used were co-occurrence network and hierarchical cluster analysis for macro-level visualization [a,p], and code similarity matrix, word association and keyword-in-context analysis for exploring semantic contexts [s,p].

An examination of the headlines to get an impression of the main themes in the corpus (cf. Machin and van Leeuwen, 2007: 9) shows a low frequency of the main keywords *LAWS (30 per cent) and *KillerRobot (13 per cent) in the headlines of the articles. This indicates a limited initial visibility of the topic in the newspapers and a discussion of LAWS in some other larger context, such as AI, but no major thematic trend could be identified.

A deeper exploration of the article texts using keyword (that is, statistically salient tokens) clustering [p,a] and contextual analysis [p,s] reveals themes in the form of statistically salient patterns, which can be seen as social action crystallized in language use (Bubendorfer, 2009: 161) – the very fabric of discourse. LAWS were found to be discussed in the context of several distinct topics, which can broadly be divided into two groups: LAWS in the context of policy and regulation; and reflections on the current and future role of autonomous AI and robots in armed conflict.

While factual elements of the discourse layers examined in earlier dominate the former area of topics, some striking examples of sociocultural imaginaries and symbolic language can be found in the latter. However, it should be noted that articles often cover several topics at different levels of detail; noticeable correlations between topics are expanded upon where deemed relevant.

[23] [s]entence, [p]aragraph, [a]rticle; [n]ouns], [ad]jectives], [v]erbs.

3.1 Newspaper coverage of lethal autonomous weapon systems policy and regulatory efforts

The UN and/or CCW/GGE are mentioned in 43 per cent of all articles, making talks and negotiations concerning LAWS regulations one of the most visible topics. The first GGE closed-door meeting in 2013 acted as a kickstart for journalistic LAWS discourse as the subject of around 50 per cent of all LAWS-related articles between 2013 and 2014, and while coverage drops significantly afterwards, CCW/GGE talks remain a main topic.

Articles that specifically mention the CCW/GGE mostly do so to give background information before diving into specific topics, with only some 7 per cent of all articles solely focused on conveying information about CCW/GGE, its work and achievements. Half of all instances [p] mentioning the CCW/GGE focus on Japanese participation; this and the relaying of its stance, mostly through partial quotes, constitute the main way in which the Japanese government is framed as active.[24] While the government's position and reasoning on LAWS are mostly passed on factually without much comment, this does not mean that there is no criticism: some 15 per cent of the corpus consist of editorials, which serve as a more direct way of expressing a publication's stance than topic, information or actor selection. That being said, only a handful of editorials call for a ban; however, many share a muted criticism of a government that is seen as too passive, and a generally critical attitude towards LAWS, which is visible in the demand for strict regulations and the description of potential threats such as warfare spiralling out of control, civilian killings or ethical dilemmas. The prevalence of verbs or phrases [p] indicating deontic modality[25] in these editorials is striking, in particular a call to action to the international community, and especially to the Japanese government, which is urged to take a more active, leading role in international regulatory efforts:

> The impression is undeniable that the Japanese government has been too timid. Appropriate regulation and control of AI-based weapons in general will be necessary in the future. (AS 2019a)

> Japan takes the position that it will not develop fully autonomous weapon systems. Hopefully, the government will play a leading role in constructive discussions. (YS 2019b)

[24] Defined here as the subject of a transitive verb indicative of action.

[25] Necessity; in this context corresponding to English have to, must, should, ought to, had better and so on.

In the context of international regulation efforts, *NGO was found to be distributed rather evenly, with 19 per cent of the corpus containing the code, and *NGO mainly being associated with participating in CCW/GGE efforts and advocating a ban or regulation of LAWS. However, NGOs are mostly referred to by this generic term and there is no coverage of Japanese anti-LAWS NGOs, their events or the information material they publish in Japan. The few NGOs that are referred to by name are international NGOs that are briefly mentioned with their respective positions, indicating a tendency to frame NGOs as 'external'.

This is different from *Experts,[26] another group of discourse actors found in 44 per cent of articles featuring CCW/GGE. *Experts are actively invited to provide commentary on CCW/GGE efforts, and to discuss topics such as ethics and drone warfare. Notable examples include robotics researcher Hirose Shigeru, astrophysicist and noted science writer Ikeuchi Satoru, and political scientist Satō Heigo; the tone of *Expert commentary varies from balanced commentary to outspoken criticism.

For example, as the only Japanese scientist to sign an international open letter in 2017 calling for a ban on the development of LAWs, Hirose is interviewed several times. When asked why he signed the letter in an interview, he responded: 'Not so long ago, machines capable of killing people without human intervention were something of the realm of science fiction. But with the rapid progress of AI technology, this could become reality. When I realised this, I felt a sense of crisis' (AS, 2020). This is also notable as an example of the use of (science) fiction as a rhetorical device (see later on in the chapter) and incidentally the only article in the corpus to discuss the issue of Japan's Constitution, with Hirose echoing the SCJ's misgivings about military research and warning that 'as long as Japan has a peace constitution, certain boundaries must not be crossed'.

Ikeuchi, perhaps unsurprisingly (see earlier in the chapter), is much more blunt in his scepticism. In an article entitled '"Drones are cruel weapons that have no remorse" warns Ikeuchi, Professor Emeritus at Nagoya University', he pulls apart many of the arguments for the development of LAWS heard in the discourse, such as that AI-powered weapons could lead to more ethical warfare:

> [I]t is clear that all weapons are designed to kill people, and there is no such thing as an ethical weapon. [LAWS] are talked about as 'weapons to save soldiers' lives', but to me that is just black humour … Instead of mulling over the ethicality of arms, we should pursue a path of no armaments. (AS 2022b)

[26] Any individual identified as scholar or researcher, or specifically labelled 'expert' in an article was included in *Experts (20 per cent of the corpus).

In comparison, Satō Heigo, the most quoted *Expert in the corpus, uses a much more pragmatic tone. Satō is often quoted as stressing the importance of the GGE talks, which he sees as having no alternative: 'It may be extremely difficult to create norms in a climate of mutual mistrust, but abandoning the regulation of AI weapons is just too dangerous' (YS 2022).

3.2 Current and future involvement of autonomous artificial intelligence in warfare and national security

In the discussion of current (semi-)autonomous weapons and speculation about the future of LAWS and autonomous AI, fact and what-if scenarios become interwoven. While part of the discussion focuses on tangible issues such as the status quo of drone and AI warfare, uncertainty and fear also crystallize in some striking examples of social imaginaries, in the sense of arguments based on what Jürgen Link calls collective symbols: '[T]he whole "imagery" of a culture, the whole of its widely used allegories and emblems, metaphors, examples, illustratives, guiding themes, comparisons and analogies' (Link, 1996: 25). The following subsections will examine the most prominent themes and examples of symbolic language in the corpus.

3.2.1 Lethal autonomous weapon systems and ethics

As the question of an AI deciding the fate of a human being is central to the discourse on autonomous AI and LAWS, it is unsurprising that *rinri* (ethics) features prominently (26 per cent[a][27]) in the corpus. A closer look reveal *Ethics to be closely related to *IHL, and described as one of the main issues of international regulatory efforts, as in 'International rules and codes of ethics for the operation of attack drones, including LAWS, are not yet established' (AS 2022a). This example also illustrates the imprecision of nomenclature that can be observed regularly; 'drones' in a headline may be referred to as 'LAWS' in the body of the article. Similarly, most articles do not bother to qualify the meaning of 'autonomous', sometimes using 'fully autonomous' or 'autonomy with some human involvement' instead when quoting the government or other sources, which ultimately points to yet unresolved issues of definitions inherited from the larger policy and technical discourses.

Predictably, the CCW/GGE features prominently in the context of *Ethics. Its purpose is explained as regulating 'inhumane weapons', implying that LAWS could also fall into this category. Apart from this association with regulatory efforts, *Ethics returns many unspecific or abstract instances

[27] See fn 23.

such as: 'It is a question that is deeply concerned with human existence and ethics: What is war without a shred of humanity, and will society tolerate it?' (AS 2019b). This example is also indicative of a trend that will be explored in the following two subsections: the framing of LAWS as the emotionless, and therefore dangerous, part of a machine-human dichotomy.

3.2.2 Symbolic representations: killer robots and the end of the world

The expression of fears and worries regarding LAWS and autonomous AI is another distinctive aspect of the discourse. Translated into a code, *Worries is present in about one third of all articles, sometimes offset by possible positive sides of using autonomous systems, the most frequent actually being nonmilitary use (disaster relief, robotics for nursing care); however, these points are usually raised in the context of the Japanese government's position on LAWS and therefore mostly provided as uncommented information.

As for specific fears expressed in the corpus, on closer examination [p,s], a number of issues emerge as the main objects of concern:

- proliferation of LAWS to terrorists or authoritarian regimes;
- fear of unmitigated bloodshed;
- loss of control;
- 'killer robots'.

The first point can be found regularly, but it is usually left without comment as part of a list of common concerns; the second issue is encapsulated in keywords such as civilian casualties, indiscriminate killing, escalation of conflicts and lowering of inhibitions to go to war. Going deeper, these concerns are often explained in terms of the final two points on the list, which warrant more elaboration.

'Loss of control' is circumscribed by various keywords relating to AI, such as hacking, malfunction, rogue AIs, 'black box phenomenon'[28] or AI singularity.[29] This fear of humans losing control of their own creations is pushed to its extreme in the eschatological image of the 'end of humanity', as in the following examples:

> If the nations of the world start to use weapons that kill at their own whim, the destruction of the human race seems closer than ever. (MS 2013)

[28] 'Black box' describes an AI whose inputs and inner workings are opaque to observers.

[29] A term used to describe the hypothetical point at which an AI surpasses human intelligence.

> Can we coexist with an evolving AI – Would it destroy humanity? (AS 2016)

A similar symbol can be found in the image of revolution (*kakumei*); 11 per cent of all articles describe LAWS as the '3rd military revolution' after gunpowder and nuclear weapons. Revolution is as ambiguous an image in Japanese as it is in English; it describes a disruption of the familiar order of things, and evokes visions of new and exciting possibilities as well as turmoil and strife. In a subtle way, it also carries the image of being put on the spot: if a revolution is seen as unstoppable, it might be better to roll with it. In a similar way to 'killer robot' (which will be discussed subsequently), it seems that this image was imported into the Japanese discourse, probably through coverage of the Heyns Report (Heyns, 2013), which uses the phrasing.[30]

The 'third revolution' is a recurring theme in all the discourse media examined, although the context suggests divergent readings. While NGOs, for example, tend to view this revolution with apprehension, the government seems to see it as a development that Japan cannot afford to miss. This is also hinted at by the fact that LAWS technology is sometimes referred to as a 'gamechanger' in the policy discourse.

The last item on the list, 'killer robot' (*satsujin robotto*) is doubtlessly the most salient example of symbolic language in the discourse, and found in almost half of the corpus [a]. Conjuring up an image of threat to life and to humanity, and of lack of empathy without the need to elaborate, 'killer robot' serves as a good example of a collective symbol, and the mere frequency of the term may serve as an indicator for an overall critical discourse. This is corroborated by the fact that in 75 per cent of instances [s], ⋆KillerRobot is not put in brackets, quotation marks and so on to suggest a quotation, a neologism or a deliberately allegorical use of language.

While the term itself is suggestive, its associations are often made explicit when LAWS are described as 'mindless killing robots, killing without regard to who gets hurt' (MS 2013), or as having 'neither hesitation to kill nor fear of being killed' (AS 2021). The image of the 'killer robot' is based on the tacit assumption of *positive* emotions (for example, compassion), and while both the mention of a lack of emotions and the aspect of AI weapons free from human error are found in the corpus, there is only one instance where the two are actually contrasted: 'Some argue that these weapons, which are also referred to as killer robots, are more accurate in identifying and attacking the enemy and thus reducing non-combatant casualties, precisely because they are not bound by emotions or physical condition'(AS 2021).

30 Ultimately pointing to Peter Singer's *Wired for War* (2009) as a source (Heyns, 2013: 5).

Apart from this instance, negative emotions that may play a role in armed conflict, such as hatred or emotionally charged biases like racism, are not explicitly mentioned in this context, and nor is the problem that an 'emotionless' AI may well inherit unreflected biases from its programmers or datasets discussed.

While 'killer robot' is virtually absent from official documents, the term features prominently in material produced by Japanese and international NGOs such as the Campaign to Stop Killer Robots. It has generally been noted to be associated with international NGOs (Fukui, 2017: 155; Shinpo, 2020: 227), and likely found its way into the Japanese LAWS discourse from English through Japanese translations of NGO material. Outside of the LAWs discourse, the phrase was used as early as the 1980s predominantly in the context of comics and science-fiction films to describe 'evil robots'. As the following section will explore, the corpus reflects this connection to fiction in Japanese to a certain degree; digging deeper, 'killer robot' emerges as part of a multilayered network of symbols that construct a subtle image of LAWS as a thing of fiction, of the future and, to a lesser extent, as something external.

3.2.3 Lethal autonomous weapon systems between the Terminator and domestic robot narratives

Satō Heigo remarks that in the discussion of LAWS, two threat narratives are often intertwined: one realistic, centred on unmanned drone strikes, and the other fictitious, often using the homicidal antagonist of the *Terminator* film franchise as a symbolic representation of the dangers of LAWS (Satō, 2014). Indeed, *ScienceFiction can be found in roughly one third of all articles, with Terminator as the most distinct correlator (39 per cent of all code instances), followed by two well-known manga protagonists, Tetsuwan Atomu and Doraemon, and US science-fiction writer Isaac Asimov. Asimov is mentioned as the author of the 'Three Laws of Robotics'; another set of similar laws by manga author Tezuka Osamu is also found. Although both 'laws' were created in the context of fiction, it is easy to see why they are used to illustrate the need for humanistic rules, given the lack of real ones, perhaps also expressing a desire for clear guidelines to distinguish 'good' from 'bad' robots or AI. As a symbol for something that is hard to perceive as reality and to reconcile with one's vision of the real world, science fiction is occasionally used to create a contrast with reality:

> Advanced robots threaten humans – a theme that has become a 'staple' of science fiction films. But robot technology is about to enter our lives in the real world. (AS 2014)

> The time may come when such weapons will be used in reality, as they are now in science fiction films. (AS 2018a)

These examples also illustrate another layer of symbolism: by framing LAWS as science fiction, they are also constructed as not yet existing, but with the implicit threat of becoming reality at some point; this is also visible in the strong correlation [p] of *NotYet with *LAWS, but also *ScienceFiction.

Incidentally, the use of science fiction as a symbol is not limited to the newspaper corpus; it is also used by NGOs and even in the political discourse – for example, when law maker Asakawa Yoshiharu specified LAWS in 2022 as 'fully autonomous, like those you often see in SF movies', adding that 'In SF movies and books, these things often end up dominating the humans' (NDL, 2022).

Of all cultural symbols in the corpus, the Terminator is certainly the most salient one; Arnold Schwarzenegger's iconic performance of an emotionless killer robot from the future has had a lasting influence on the perception of robots in general, and the discussion of AWS in particular. Used mostly as an uncommented allegory, the Terminator is a picture-book example of a self-explanatory collective symbol for AI/robotics-related threats, and as such is not limited to Japan. In his book *Army of None*, Paul Scharre notes that he was '[c]ontinually struck by how much the Terminator films influence the debate on autonomous weapons. In nine out of ten serious conversations on autonomous weapons I have had, whether in the bowels of the Pentagon or the halls of the United Nations, someone invariably mentions the Terminator' (Scharre, 2018: 264).

This tendency is visible in the corpus as well; representative examples include a 2018 Asahi headline: 'Is the world of Terminator becoming reality? When AI becomes a killer weapon' (AS 2018b) or 'LAWS have also been called "killer robots" by the media. What naturally comes to mind here is the Terminator from the movie' (YS 2019a).

Similar to the more general use of 'science fiction' detailed earlier, the symbolic power of the Terminator is not limited to a specific field of discourse; it appears in the academic (Akimoto, 2019) as well as in the political discourse, as the example of Foreign Minister Kōno using the image to illustrate the risks of AI weapons shows (MOFA, 2019b), which also substantiates Scharre's remark. However, the liberal use of the Terminator symbolism is not without criticism; Satō calls it 'too unrealistic and only stirring up fear' (AS 2018a), a sentiment echoed by Tomikawa Hideo, a researcher at the National Defence Research Institute, stating in an interview that '[People] tend to think that AI can make anything possible, but it is a pipe dream to think that there will be a "Terminator"' (MS 2023).

Upon closer examination of Terminator [p], a partly connected additional symbolic layer emerges, best described as 'Terminator-Atomu dichotomy'. While the manga character Tetsuwan Atomu, a robot in the entirely unthreatening looking form of a boy (see Figure 3), has enjoyed success as Astro Boy outside of Japan, the character is used in all instances in the

Figure 3: Tetsuwan Atomu/Astro Boy

Source: LostplanetKD73 (2019)

corpus as a specifically Japanese symbol, just as the Terminator is the fictitious Other; a comparison that is not limited to the LAWS discourse (cf. Allgaier, 2014: 37).

This image is aptly expressed in a 2003 opinion piece on (then hypothetical) combat robots: 'Atomu would probably scream "Don't use robots like that!"' (AS 2003). The 'friendly robot' symbolism is also linked to the image of a general Japanese fondness for robots, extending the Terminator-Atomu juxtaposition to a contrast between Japanese and 'Western' culture, as in these examples:

> What is interesting is that these films are coming out of the West. In films such as 2001: A Space Odyssey and Transcendence, artificial intelligence and human beings are usually depicted as enemies. In Japan, on the other hand, there are many films that depict robots as friends, such as Tetsuwan Atomu and Doraemon. (YS 2015)

> In Japan, AI transcending human intelligence tends to be portrayed as a human-like character like Doraemon, whereas in other countries it is more of an inhuman character like the Terminator. (YS 2021)

The Japanese self-image as a robot-friendly society is in fact thought to have arisen from the huge popularity of Tetsuwan Atomu in the postwar era (Allgaier, 2014: 67), but while there is little evidence for the existence of such a positive bias in real life (MacDorman et al, 2009: 507; Prochaska-Meyer and Wagner, 2022: 85),[31] the image of a general fondness for robots exists as an internalized image as well as an external one. For example, Scharre remarks that in 'Japanese culture, robots are often seen as protectors and saviors. Some people see autonomous weapons as inherently wrong. Others don't' (2018: 307), suggesting a greater potential for acceptance of LAWS in Japan than in other cultures.

In the context of LAWS, the Japanese government stresses military-civilian dual use of autonomous robot technology. One use-case in particular that is frequently mentioned is robots to offset the demographic problem of a rapidly ageing society. *Kaigo* (nursing care) and *fukushi* (social welfare) are the key terms of this topic, which can be found in 7 per cent of the articles. While robotic technology for use in nursing homes has long featured prominently in Japanese plans for a robotized society, and in the early 2010s was considered an emerging key sector in robotics (Hook et al, 2015: 84), the image does not match the reality; the practical implementation of such technology remains very limited, and robots have generally proved unpopular with care workers (Prochaska-Meyer and Wagner, 2022), leaving little to show a decade later.

However, the discursive power of a collective symbol lies precisely in the fact that neither explanation nor evidence is needed; it is the symbolic version of a circular argument, as it were.

3.2.4 The Russo-Ukrainian war: a discursive reality check?

When politics researcher Enomoto Tamara predicted in an interview in 2020 that the fictitious facets of the discourse would get gradually replaced

[31] See Wagner, C. (2013) *Robotopia Nipponica – Recherchen zur Akzeptanz von Robotern in Japan*, Marburg: Tectum, for a detailed discussion.

by 'a sober and down-to-earth discussion on existing armed drones' (MS 2020), she probably did not imagine it would happen quite so fast. With the Russian invasion of Ukraine in 2022, reports on 'kamikaze drones' and similar types of semi-autonomous weapons have become an almost daily occurrence; this is also visible in the journalistic discourse on LAWS: In the 12 months since the start of the war in March 2022, 22 articles were published discussing LAWS in the larger context of the war in Ukraine, which amounts to about 13 per cent of the corpus.

While these articles mostly deal with LAWS only in a subsection of the text, they nevertheless serve to offset the symbolism-driven scenarios described previously; many focus on drone warfare, leading to a significant spike in this topic for 2022–2023. The war in Ukraine is described as a new kind of war, zooming in on weapons such as loitering munitions, as well as the use of social media and IT warfare. *LAWS are mostly focused on towards the end of an article, when the stance of the Japanese government on LAWS is discussed, and future developments extrapolating from recent ones are speculated upon. An *Asahi* article dating from March 2023 encapsulates this emerging connection of weapons in actual use with a new sense of urgency of dealing with the issues of LAWS: 'The international community must stop Russia's acts of aggression as soon as possible. At the same time, there is a need to urgently create international norms that respond to the reality of "new wars" created by the evolution of information technology' (AS 2023).

The war in Ukraine also seems to have acted as a catalyst for the government's plans to bring the JSDF up to modern standards, including AI technology and UVs, and to generally increase defence capabilities. After the government announced its plans to increase defence spending in late 2022, several articles, focused on emerging technologies as one of the key areas mentioned in the budget proposal, were published; a connection to the war in Ukraine is visible in most of these articles, for which an *Asahi* piece entitled '"Almost like a trade fair": the war in Ukraine has moved the Ministry of Defence to make drones a mainstay' is a good example. The article gives a detailed report on the background of this development, including the initially limited willingness of the MOD to adopt drones because of high research and development costs. However, this seems to have changed with the war in Ukraine (AS 2022a), and in a similar piece from October 2022, *Mainichi* zooms in on the JSDF's plans to purchase combat drones for the first time, quoting Hideaki Watanabe, First Director-General of the Defence Equipment Agency: 'It would bring us several steps closer to a "combat-ready Self-Defence Force"' (MS 2022).

Incidentally, this article is also notable for its challenge of the government's pledge not to introduce fully autonomous weapon systems, when an 'anonymous MOD official' is quoted as questioning whether Japan could afford to limit its defence capabilities based on such self-imposed ethical norms when countries like China would do no such thing (AS 2022a), an argument that was also encountered in parliamentary debates (see the earlier section on political parties) and sheds some light on the internal political debate.

Conclusion

The public discourse on LAWS in Japan spans a decade of lively debate and several interwoven layers: researchers collaborate with NGOs actively campaigning for LAWS awareness in Japan; both appear as advisors in government discussions and as members of working groups; the government discusses policy in national and international contexts; and newspapers report the status quo of policy discourse and discuss future developments while offering a platform for researchers and other experts to provide commentary. While it can be challenging to pinpoint the effect of these interactions, a clear example of the interconnectedness of discourse layers and actors is the call from NGOs, newspapers and law makers for the Japanese government to pursue an active role in international regulatory efforts, reflecting an expectation for Japan to spearhead such efforts. This seems to build on the premise that Japan has an obligation to 'lead the discourse as the only nation that has made a clear decision on how not to use LAWS', as law maker Miura Nobuhiro (NKP) stated in a session of the Committee on Foreign Affairs and Defence (NDL, 2021); similar expectations can be found in NGO material and newspaper articles. However, for the most part, this 'obligation' is not explained and can be seen as a manifestation of the lingering anti-militarism that has informed the Japanese security identity for many decades.

Against this backdrop, the Japanese government has adopted a pragmatic approach to LAWS, leaving the door open to the potential use of LAWS in a broader sense, and indeed recently starting to adopt LAWS technology as part of its plans to strengthen military capabilities, while simultaneously taking great pains to assuage the 'allergy to everything to do with the military' (Iwamoto, 2021: 2) in Japanese society. This duality is reflected in the government's framing of LAWS not primarily in terms of military benefits, but in terms of civilian applicability and functional necessity, thus integrating LAWS into the still-prevailing peace state master narrative via a rhetoric of inevitability in the face of internal and external developments, as well as by emphasizing the benefits of autonomous robot technology for society at large. In the context of international regulatory efforts, the Japanese government has presented itself as a staunch supporter of LAWS regulation, distinguishing between fully autonomous human-out-of-the-loop systems considered incompatible with IHL, and semi-autonomous systems with a 'meaningful degree of human interaction', which are not considered true LAWS. Japan's adoption of a distinction similar to the US concept of LAWS in the GGE arguably also reflects its strategic alignment with the US in the context of the US-Japan Security Treaty. While the US and Japan can look back on over 70 years of cooperation under the umbrella of the US-Japan Security Treaty, it is hard to overlook that this cooperation has never been on an equal footing. Ultimately, potential US military assistance in the event of an attack on Japan is an integral part of Japan's defence strategy, not the other way round. It is therefore not surprising that efforts are being made to balance the desire to maintain a consistent pacifist (self-)image with the need not to antagonize Japan's closest military ally. However, as the US definition of autonomous systems is changing to potentially encompass systems with nonhuman operators (cf. Valadares Fernandes Barbosa, 2023), it will be interesting to see if Japan will adapt its interpretation of LAWS in its approach to international and domestic discourse.

Overall, it seems safe to assume that the public discourse on LAWS in Japan has parallels to similar discourses in other countries, such as focusing on ethical considerations and technological feasibility, and highlighting the destructive and disruptive potential of autonomous weapons. However, sociocultural factors such as the aforementioned security identity and collective symbols have also shaped the discourse, such as the use of fictional characters to represent contrasting archetypes of robots: 'evil' and foreign (the Terminator) versus 'good' and domestic (Tetsuwan Atomu). Through such symbolic language, media and political discourse is encoded in intuitively understandable symbols, and in the case of LAWS, the various discourse layers were found to use imaginaries like the Terminator, the term 'killer robots' or generally science fiction as stand-ins for a rather intangible threat with little connection to subjective reality. However, as a result of the war in Ukraine, there is now a tendency for these imagined LAWS to be replaced by actual unmanned vehicles or drones, which now have model and manufacturer names attached, and whose use is regularly reported in the newspapers; additionally, the increasing ubiquity and thus normalization of AI in public discourse with the popularization of LLMs since 2022 may also have contributed to desensitizing the public to the idea of AI-controlled robots in general. Moreover, the Russian invasion of Ukraine has also provided the government with a new external threat scenario to rationalize and accelerate the already ongoing upgrading of military capabilities (Okuno, 2023: 138), including unmanned vehicles, AI and other emergent technologies (MOD, 2022). As the security identity of domestic antimilitarism is thus increasingly replaced by a new narrative of a more assertive 'proactive peace state', as domestic opposition to changing the Peace Constitution wanes among younger people (HSB, 2023; YUSO, 2023), and given that the Japanese government has not signed any binding document or passed any legislation on the use of LAWS, Japan's stance on LAWS may therefore well change in the long term. It remains to be seen how such a development would resonate in the public discourse.

References

AAR Japan (Association for Aid and Relief Japan) (2013) 'Kōenkai 'Satsujin robotto no kinshi ni mukete' 11/08 kaisai – Nippon umare no kokusai NGO AAR Japan nanmin o tasukeru kai' ['Towards a ban on killer robots' lecture], *AAR Japan*, 8 November. Available from: https://aarjapan.gr.jp/about/news/2013/1024_1360.html [Accessed 25 July 2022].

AAR Japan (2014) *We don't need 'killer robots' in this world!* Available from: https://aarja pan.gr.jp/about/news/ images/news1603_20141robot.jpg [Accessed 23 January 2025].

AAR Japan (2016) 'Kirā robotto ni tsuite manaberu bukkuretto o goriyō kudasai' [Use our booklet to learn about killer robots]. Available from: https://aarjapan.gr.jp/about/news/2016/0401_2014.html [Accessed 17 July 2022].

AAR Japan (2018) '2. Kirā robotto no nai sekai ni muketa Nippon no yakuwari o kangaeru benkyōkai' [Second study group on Japan's role in a world without killer robots].Available from: https://www.aarjapan.gr.jp/about/news/docs/press1811_2660_killer.pdf [Accessed 17 July 2022].

Akimoto, D. (2019) 'International regulation of "lethal autonomous weapons systems" (LAWS): paradigms of policy debate in Japan', *Asian Journal of Peacebuilding*, 7(2):311–332.

Allgaier, A. (2014) 'Nihon-robotto-ron: Eine Dekonstruktion der These von der japanischen Robotophilie' [Nihon-robotto-ron: Deconstructing the idea of a robot-loving Japan], MA thesis, University of Vienna.

Berkofsky, A. (2010) 'Japan's post-war constitution: origins, protagonists and controversies', *Il Politico*, 75(2): 5–25.

Bosack, M. (2023) 'Kishida's proposed tax cuts put higher defense spending on hold', *Japan Times*, 23 October. Available from: https://www.japantimes.co.jp/commentary/2023/10/23/japan/kishida-tax-cuts-defense-budget/ [Accessed 11 January 2024].

Bubendorfer, N. (2009) *Sprachgebrauchsmuster: Korpuslinguistik als Methode der Diskurs- und Kulturanalyse* [Patterns of language use: corpus linguistics as a method for discourse and cultural analysis], Berlin: De Gruyter.

Cabinet Secretariat of Japan (2022) *National Security Strategy (NSS) – English Version*. Available from: https://www.cas.go.jp/jp/siryou/221216anzenhoshou/nss-e.pdf [Accessed 11 January 2024].

CSKR (Campaign to Stop Killer Robots) (2022) 'Stop killer robots'. Available from: https://www.stopkillerrobots.org/about-us/ [Accessed 25 July 2022].

DWD (Shūkan Daiamondo) (2021) '2020 nen Sōkagakkai no kinen no toshi ni roteishita saikyō kyōdan no kōzōteki kiki shūkan daiyamondo no midokoro' [2020 – the structural crisis of the strongest cult exposed in the anniversary year of the Soka Gakkai/highlights from *Diamond Weekly*], *Diamond Weekly*. Available from: https://dw.diamond.ne.jp/articles/-/29789 [Accessed 25 July 2022].

Foucault, M (2004) *The Archaeology of Knowledge*, London: Routledge.
Fukui, Y. (2017) 'Aratana gijutsu to kokusaihō no tekiyō kanōsei' [New technologies and the applicability of international law], *Sekaihō nenpō*, 36: 154–185.
Funabashi, Y. (2016) 'Japan's silver pacifism', *The National Interest*, 141: 25–31.
Gustafson, K. (2016) 'Identity and recognition: remembering and forgetting the post-war in Sino-Japanese relations', in L. Hagström (ed) *Identity Change and Foreign Policy: Japan and Its 'Others'*, New York: Routledge, pp 117–138.
Haddad, M. and Hussein, M. (2021) 'Infographic: history of US interventions in the past 70 years', *AlJazeera.com*, 10 September, Available from: https://www.aljazeera.com/news/2021/9/10/infographic-us-military-presence-around-the-world-interactive [Accessed 27 July 2022].
Hatakeyama, S. and Kawasaki, T. (2018) *Manga Nyūmon Satsujin Robotto Ga Yatte Kuru?!* [Manga Primer: Are the Killer Robots Coming?!], Tokyo: Gōdō shuppan.
Heyns, C. (2013) 'UN DOC A/HRC/23/47 = Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Christof Heyns'. Available from: https://digitallibrary.un.org/record/755741 [Accessed 18 July 2022].
Hook, G.D., Mason, R. and O'Shea, P. (2015) *Regional Risk and Security in Japan: Whither the Everyday*, New York: Routledge.
HSB (*Hokkaidō shinbun*) (2023) '"Sensō sedai" genshō eikyō ka kyū jō 'kaiseisu beki' saita ni honshi zen michi yoron chōsa' [Impact of the dwindling numbers of the 'war generation'? The number of people who think Article 9 'should be revised' is at a high – Hokkaidō public opinion poll], 27 April. Available from: https://www.hokkaido-np.co.jp/article/838615 [Accessed 10 June 2023].
Hughes, C.W. (2011) 'The slow death of Japanese techno-nationalism? Emerging comparative lessons for China's defense production', *Journal of Strategic Studies*, 34(3): 451–479. https://doi.org/10.1080/01402390.2011.574987
Hughes, C.W., Patalano, A. and Ward, R. (2021) 'Japan's grand strategy: the Abe Era and its aftermath', *Survival*, 63(1):125–160. https://doi.org/10.1080/00396338.2021.1881258 [Accessed 10 June 2023].
International Committee of the Red Cross (2022) 'Jiritsu heiki shisutemu o meguru IRC no kenkai' [ICRC's statement regarding AWS]. Available from: https://jp.icrc.org/information/aws-position-paper/ [Accessed 20 June 2022].
Iwamoto, S. (2021) 'Jieitai to kokusaihō no kankeisei no hensen: Jiko yokusei to hōteki zure o koete' [The changing relationship between the Self-Defence Forces and international law: looking beyond self-control and legal deviation], *Sandai hōgaku*, 55(2): 213–282. Available from: https://ksu.repo.nii.ac.jp/records/10656 [Accessed 10 June 2023].

Iwamoto, S. (2022) 'Jiritsugata chishi heiki shisutemu LAWS: Kisei tasōteki takakuteki apurōchi kara no kisei no kokoromi – Tokushū: Shin gijutsu dōkō to gunshuku' [Regulation of autonomous lethal weapon systems (LAWS) – a multilayered and mult-faceted approach to regulation], *Disarmament Review*, 11(1): 5–22.

Jäger, S. (2001) 'Discourse and knowledge: theoretical and methodological aspects of a critical discourse and dispositive analysis', in R. Wodak and M. Meyer (eds) *Methods of Critical Discourse Studies*, London: Sage, pp 32–62.

Jäger, S. (2015) *Kritische Diskursanalyse. Eine Einführung* [Critical Discourse Analysis: An Introduction], 6th edn. Duisburg: Unrast (Edition DISS).

JNA (Japanese Newspaper Association) (2021) 'Shinbun no sesshoku jōkyō' [Newspaper exposure]. Available from: https://www.pressnet.or.jp/adarc/data/read/data01.html [Accessed 22 July 2022].

JNA (2021) 'Shinbun no sesshoku jōkyō' [Status of exposure to newspapers]. Available from: https://www.pressnet.or.jp/adarc/data/read/data01.html [Accessed 13 January 2024]

JNA (2023) 'Shinbun ōjiiensu chōsa teiten chōsa' [Newspaper audience survey (fixed-point survey)]. Available from: https://www.pressnet.or.jp/adarc/data/audience/report.html [Accessed 11 June 2023].

Kabata, H. (2018) 'Jinkō chinō no gunji ōyō' [Military applications of artificial intelligence], *Kagakugijutsu shakairon kenkyū*, 16: 30–42. https://doi.org/10.24646/jnlsts.16.0_30.

Kallender, P. and Hughes, C.W. (2019) 'Hiding in plain sight? Japan's militarization of space and challenges to the Yoshida Doctrine', *Asian Security*, 15(2): 180–204. https://doi.org/10.1080/14799855.2018.1439017

Kanzaki, N. (2021) 'Korekara no robotto kenkyū no tame no rinri' [Ethics for the future of robotic research], *Nihon robottogakkaishi*, 39(1):18–21. https://doi.org/10.7210/jrsj.39.18

Keller, R. (2013) 'Das wissen der wörter und diskurse' [The knowledge of words and discourse], in W. Viehöfer (ed) *Diskurs – Sprache – Wissen. Interdisziplinäre Beiträge zum Verhältnis von Sprache und Wissen in der Diskursforschung* [Discourse – Language – Knowledge. Interdisciplinary contributions on the relationship of language and knowledge in discourse studies], Wiesbaden: Springer Fachmedien, pp 21–49.

Kukita, M. et al (2014) 'Another case against killer robots', in *Sociable Robots and the Future of Social Relations*. Amsterdam: Ios Press, pp 291–295.

Landwehr, A. (2001) *Geschichte des Sagbaren. Einführung in die historische Diskursanalyse* [History of the Sayable: An Introduction to Historical Discourse Analysis], Tübingen: Edition diskord.

Link, J. (1996) *Versuch über den Normalismus. Wie Normalität produziert wird* [A Shot at Normalism. How Normality is Produced.], Opladen: Westdeutscher Verlag.

LostplanetKD73 (2019) Tetsuwan Atomu/Astro Boy. Available from: https://commons.wikime dia.org/ wiki/Category:Astro_Boy#/media/File:NYCC_2021 03.jpg (CC BY- SA 4.0) [Accessed 23 January 2025].

MacDorman, K.F., Vasudevan, S.K. and Ho, C.-C. (2009) 'Does Japan really have robot mania? Comparing attitudes by implicit and explicit measures', *AI & Society*, 23(4):485–510. https://doi.org/10.1007/s00146-008-0181-2

Machin, D. and van Leeuwen, T. (2007) *Global Media Discourse: A Critical Introduction*, Abingdon: Routledge.

Maslow, S. (2015) 'A blueprint for a strong Japan? Abe Shinzo and Japan's evolving security system', *Asian Survey*, 55: 739–765. https://doi.org/10.1525/as.2015.55.4.739

Matsuyama, N. and Sato, T. (2021) 'Japan set to pay significantly more to host US military bases', *The Asahi Shimbun*. Available from: https://www.asahi.com/ajw/articles/14487877 [Accessed 15 May 2023].

MOD (Japanese Ministry of Defence) (2014) 'Kaisetsu kakudaisuru mujinki kaihatsu no dōkō, Heisei 26 nenban bōei hakusho' [Commentary: trends in unmanned aircraft development, 2014 White Paper on Defence]. Available from: http://www.clearing.mod.go.jp/hakusho_data/2014/html/nc002000.html [Accessed 10 April 2023].

MOD (2019) 'Kaisetsu jiritsugata chishi heiki shisutemu LAWS ni tsuite, Reiwa gannenban bōei hakusho' [Commentary: autonomous lethal weapon systems (LAWS), 2019 White Paper on Defence]. Available from: https://www.mod.go.jp/j/publication/wp/wp2019/html/nc029000.html [Accessed 21 July 2022].

MOD (2020) 'Shiryō 2: Heisei 31 nendo ikō ni kakaru bōei keikaku no taikō nitsuite, Reiwa 2 nenban bōei hakusho' [Document 2: outline of defence plans for FY 2019 onwards, 2020 White Paper on Defence]. Available from: http://www.clearing.mod.go.jp/hakusho_data/2020/html/ns002000.html [Accessed 11 April 2023].

MOD (2022) 'Bōeiryoku bapponteki kyōka 'gannen' yosan' [Budget for 'first year' of the major reinforcement of defence capabilities]. Available from: https://www.mod.go.jp/j/yosan/yosan_gaiyo/2023/yosan_20221223-1.pdf [Accessed 11 April 2023].

MOFA (Japanese Ministry of Foreign Affairs) (2019a) 'Working Paper: possible outcome of 2019 Group of Governmental Experts and future actions of international community on lethal autonomous weapons systems'. Available from: https://www.mofa.go.jp/mofaj/files/100113384.pdf [Accessed 11 April 2023].

MOFA (2019b) 'Dai 198 kai kokkai ni okeru Kōno gaimu daijin no gaikō enzetsu' [Minister of Foreign Affairs Kōno's address on foreign affairs at the 198th session of the Diet]. Available from: https://www.mofa.go.jp/mofaj/fp/pp/page3_002672.html [Accessed 31 July 2022].

MOFA (2023) Tokutei tsūjō heiki shiyō kinshi seigen jōyaku jiritsugata chishi heiki shisutemu nikansuru seifu senmonka kaigō no kaisai [Convention on the Prohibition and Restriction of the Use of Certain Conventional Weapons – governmental expert meetings on autonomous lethal weapon systems]. Available from: https://www.mofa.go.jp/mofaj/dns/ca/page22_004049.html [Accessed 10 April 2023].

NDL [National Diet Library] (2019) *Dai 198 kai kokkai Sangiin gaikō bōei iinkai dai 17 gō* [*198th National Assembly Senate Committee on Foreign Affairs and Defence No. 17*]. Available from: https://kokkai.ndl.go.jp/#/detail?minId=119813950X01720190613&spk [Accessed 11 January 2024]

NDL (2021) *Dai 204 kai kokkai Sangiin gaikō bōei iinkai dai 16 gō* [*204th National Assembly Senate Committee on Foreign Affairs and Defence No. 16*]. Available from: https://kokkai.ndl.go.jp/#/detail?minId=120413950X01620210603&spk [Accessed 14 April 2023].

NDL [National Diet Library] (2022) *Dai 210 kai kokkai Shūgiin anzen hoshō iinkai dai 3 gō* [*210th Session of the National Diet, House of Representatives, Committee on Security, No. 3*]. Available from: https://kokkai.ndl.go.jp/#/detail?minId=121003815X00320221027&spk [Accessed 14 April 2023].

Nihon keizai shinbun (2022) '"Kamikaze dorōn to yobanaide" Bei dantai ga' ['Please don't call them kamikaze drones', says US organization]. Available from: https://www.nikkei.com/article/DGXZQOGN2331R0T21C22A0000000/ [Accessed 25 May 2023].

NKP = Kōmeitō (2018) [NKP] 'AI heiki no kisei kakasezu' [Regulations on AI weapons are essential], *Kōmei nyūsu*. Available from: https://www.komei.or.jp/news/detail/20180413_27922 [Accessed 11 July 2022].

NKP = Kōmeitō (2019) [NKP] 'AI heiki kokusai kisei o teian' [A proposal for international regulation of AI weapons], *Kōmei nyūsu*. Available from: https://www.komei.or.jp/komeinews/p25211/ [Accessed 11 July 2022].

NKP = Kōmeitō (n.d.) [NKP] 'Heiwa no tō' ga kinkanban ['Party of peace' – that is our gold standard]. Available from: https://www.komei.or.jp/content/heiwa2018/ [Accessed 25 July 2022].

Okuno, T. (2023) '"2022 nen kokka anzen hoshō senryaku" to Nipponkoku kenpō' [The 2022 'National Security Strategy' and the Japanese Constitution], *Shakaikagaku kenkyū nenpō*, 52:1 37–147. Available from: http://hdl.handle.net/10519/10721 [Accessed 25 May 2023].

Oros, A.L. (2015) 'International and domestic challenges to Japan's postwar security identity: "norm constructivism" and Japan's new "proactive pacifism"', in L. Hagstrom (ed) *Identity Change and Foreign Policy: Japan and Its 'Others'*, Abingdon: Taylor & Francis, pp 139–160.

Oros, A.L. (2017) *Japan's Security Renaissance: New Policies and Politics for the Twenty-First Century*, New York: Columbia University Press.

PB (Peace Boat Japan) (2018) '"Manga nyūmon satsujin robotto ga yattekuru!?" Pīsubōto sutaffu ga gunji dorōn to robotto heiki no nyūmonsho o shuppan shimashita' ['Manga introduction: the killer robots are coming!' Peace Boat staff published a primer on military drones and robotic weapons. Available from: https://peaceboat.org/22359.html [Accessed 25 July 2022].

Prochaska-Meyer, I. and Wagner, C. (2022) 'Japan auf dem Weg zu einer Roboter-Gesellschaft? Visionen und Realitäten am Beispiel der Pflegerobotik', *MINIKOMI: Austrian Journal of Japanese Studies*, 89: 78–89. Available from: https://doi.org/10.25365/aaj-2022-89-07.

Reuters (2019) 'Trump says "unfair" defense treaty with Japan needs to be changed', 29 June. Available from: https://www.reuters.com/article/us-g20-japan-trump-idUSKCN1TU0AJ [Accessed 15 May 2023].

Reuters (2023) 'Exclusive: weak yen forces Japan to shrink historic military spending plan', 11 November. Available from: https://www.reuters.com/markets/currencies/weak-yen-forces-japan-shrink-historic-military-spending-plan-2023-11-03/ [Accessed 11 January 2024].

Sakurai, Y. (2020) 'Nippon gunshuku gakkai CCW – LAWS – GGE ni mukete 2019–2020 no kokusai shakai no giron to ronten seiri' [The Japanese Society for Disarmament Studies: towards the CCW-LAWS-GGE debate and issues from the international community for 2019–2020]. Available from: http://www.disarmament.jp/pdf/2020/sakurai_CCW-LAWS-GGE.pdf [Accessed 15 May 2023].

Sano, T. (2014) 'Statement by H.E. Ambassador Toshio SANO, Permanent Representative of Japan to the Conference on Disarmament, Experts Meeting of Lethal Autonomous Weapons Systems to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects'. Available from: https://www.disarm.emb-japan.go.jp/Statements/140513%20CCW%20LAWS.htm [Accessed 18 July 2022].

Satō, H. (2014) 'Chishisei jiritsu heiki shisutemu LAWS o meguru shomondai' [Issues surrounding lethal autonomous weapon systems (LAWS)], *Kokusai anzen hoshō*, 42(2): 1–14. https://doi.org/10.57292/kokusaianzenhosho.42.2_1

Satō, H. (2020) 'Jiritsugata sasshō heiki LAWS kaihatsu kinshi ni muketa ngo no yakuwari' [The role of NGOs in efforts to ban the development of lethal autonomous weapons (LAWS)], *Shakai dezaingakkai gakkaishi*, 11: 45–53. https://doi.org/10.32240/socialdesign.11.0_45

Sawa, Y. and Saisho, R. (2022) 'Japan', Reuters Institute for the Study of Journalism. Available from: https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2022/japan [Accessed 11 June 2023].

Scharre, P. (2018) *Army of None: Autonomous Weapons and the Future of War*, New York: Norton.

SCJ (Science Council of Japan) (2022) 'Science Council of Japan'. Available from: https://www.scj.go.jp/en/ [Accessed 28 July 2022].
Seligman, L. and Gramer, R. (2019) 'Trump asks Tokyo to quadruple payments for US troops in Japan', *Foreign Policy*, 15 November, Available from: https://foreignpolicy.com/2019/11/15/trump-asks-tokyo-quadruple-payments-us-troops-japan/ [Accessed 15 May 2023].
Shinpo, F. (2017) 'Robottohō o meguru hōryōikibetsu kadai no chōkan' [A bird's-eye view of issues in different areas of law concerning robot law], *Jōhō hōsei kenkyū*, 1: 64–77. Available from: https://doi.org/10.32235/alis.1.0_64 [Accessed 15 May 2023].
Shinpo, F. (2020) 'Jiritsugata chishi heiki shisutemu LAWS ni kansuru robotto hōteki shiten kara no kōsatsu' [A robot law perspective on autonomous lethal weapon systems (LAWS)], *IEICE Fundamentals and Boundaries Society Review*, 13(3): 217–30. Available from: https://doi.org/10.1587/essfr.13.3_217.
Shio (2016) 'Taidan: Yotō-Kōmeitō no shimei to sekinin o tou' [Conversation: questioning the mission and responsibility of the ruling party and the Kōmeitō], *Shio WEB*. Available from: http://www.usio-mg.co.jp/archives/1860/5 [Accessed 28 July 2022].
Shoko-Ranking (2020) 'Shinbun hakkō busū ranking: Yomiuri shinbun ga sekai ichii'[Newspaper circulation ranking – *Yomiuri Shinbun* is No. 1 worldwide]. Available from: https://www.shoko-ranking.com/entry/2020/04/18/新聞発行部数ランキング_ [Accessed 22 July 2022].
Singh, B. (2013) *Japan's Security Identity: From a Peace State to an International State*, Abingdon: Routledge.
Sōka gakkai (2022) 'kirā robotto X SDGS: Sshiru koto kara hajimaru watashi no ippo' [Killer Robot x SDGs – knowledge is my first step], *Sōkagakkai kōshiki saito*. Available from: https://www.sokagakkai.jp/picks/2256516.html [Accessed 11 April 2023].
Takahashi, M. (2021) 'Bunken tanbō: Kagakusha wa, naze gunji kenkyū ni te o somete waikenai ka' [Literature review: 'why scientists should not engage in military research'], *Ōsakaichiritsudaigaku daigaku kyōiku*, 18(2): 70–74. https://doi.org/info:doi/10.24544/ocu.20210617-004
Think Tank European Parliament (2020) *Japan's Ageing Society*. Available from: https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2020)659419 [Accessed 11 June 2023].
TVAC [Tōkyō Voluntary Action Center] (2018) 'Kirā robotto no nai sekai ni muketa Nippon no yakuwari o kangaeru benkyōkai' [Study group on Japan's role for a world without killer robots], *Borantia/Shimin katsudō no sōgōjōhō saito 'Borashiminweb'*. Available from: https://www.tvac.or.jp/sagasu/43413 [Accessed 25 July 2022].

Ueno, H. (2019) 'Robotto heiki no jiritsusei ni kan suru ichi kōsatsu: LAWS jiritsugata chishi heiki shisutemu o chūshin toshite' [A study on robotic weapons autonomy focusing on LAWS (autonomous lethal weapon systems)], *Japan Maritime Self-Defense Force Staff College Review*, 9(1):139–158.

UN Panel of Experts Established Pursuant to Security Council Resolution 1973 (2011) (2021) [UN] 'Letter dated 8 March 2021 from the Panel of Experts on Libya established pursuant to Resolution 1973 (2011) addressed to the President of the Security Council'. Available from: https://digitallibrary.un.org/record/3905159 [Accessed 25 July 2022].

Valadares Fernandes Barbosa, L. (2023) 'Exploring the 2023 U.S. Directive on Autonomy in Weapon Systems', *CEBRI Revista*. Available from: https://cebri.org/revista/en/artigo/114/exploring-the-2023-us-directive-on-autonomy-in-weapon-systems [Accessed 14 January 2024].

Van Dijk, T.A. (2001) 'Multidisciplinary CDA: a plea for diversity', in R. Wodak and M. Meyer (eds) *Methods of Critical Discourse Studies*, London: Sage, pp 95–120.

Wodak, R. (2001) 'What CDA is about – a summary of its history, important concepts and its developments', in R. Wodak and M. Meyer (eds) *Methods of Critical Discourse Studies*, London: Sage, pp 1–13.

Yoshida Y. (2020) 'Jiritsugata chishi heiki shisutemu no kisei o meguru saikin no dōkō: tokutei tsūjō heiki shiyō kinshi seigen jōyaku seifu senmonka kaigō ni okeru giron o chūshin ni' [Recent developments in the regulation of lethal autonomous weapon systems with a focus on the discussions at the Meeting of Governmental Experts of the Convention on the Prohibition of the Use of Certain Conventional Weapons], *Kokusai kōkyōseisakukenkyū*, 25(1): 23–45. https://doi.org/10.18910/77124

YUSO (*Yomiuri shinbun online*) (2023) '2023 nen 3-4 gatsu yūsō zenkoku seron chōsa "kenpō"' [March–April 2023 national opinion poll by post, 'Constitution'], *Yomiuri shinbun online* , Available from: https://www.yomiuri.co.jp/election/yoron-chosa/20230502-OYT1T50178/ [Accessed 10 June 2023].

Articles cited

Asahi shinbun

AS 2003 = 2003 04 07 'Atomu no namida wa mitakunai: Robotto sedai (shasetsu)' [I don't want to see the tears of Atomu: the robot generation (Editorial)]

AS 2014 = 2014 06 28 'Jisedai robotto hitogata, kyūjo ya kaigo ni kitai' [Next generation robots: humanoid, expected to be used for rescue and care.]

AS 2016 = 2016 04 14 'Shinka suru eiai to kyōzon suru ni wa jinrui horobosu kamo? Rinrimen no giron hajimaru' [How to coexist with evolving AI? Would it destroy humanity? … ethical debate begins.]

AS 2018a = 2018 10 20 'Satsujin robotto Shutsugen o yurushite ii ka' [Killer robots: can we allow them to emerge?]

AS 2018b = 2018 11 22 'Tāminētā no sekai ga genjitsu ni!? "AI ga satsujin heiki ni naru hi"' [Will world of Terminator become reality?! 'The day AI becomes a killer weapon']

AS 2019a = 2019 03 29 'Satsujin robotto Kōsokuryoku aru kisei mezase' [Killer robots: we need a binding regulation!]

AS 2019b = 2019 08 25 'Robotto heiki hōkisei ni mukete giron o' [Robotic weapons: discussion aiming for legal regulation]

AS 2020 = 2020 06 24 'Marude SF, Ai heiki ga mokuzen ni – Beiikoku to dōmei no Nippon wa?' [Just like in SF: AI weapons are looming on the horizon. But what about Japan and its alliance with the US?]

AS 2021 = 2021 07 05 'Satsujineiai heiki jikkō aru kisei ga kyūmu da' [Killer AI weapons: effective regulations are urgently needed] AS 2022a = 2022 11 07 '"Marude mihon'ichi": Dorōn o shuryoku e, bōeishō o ugokashita Ukurainasen ['It's almost like a trade fair': the war in Ukraine moves the Ministry of Defence to make drones a mainstay of its operations]

AS 2022b = 2022 11 10 '"Mujinki wa kōkaishinai zankoku heiki", keishō narasu Ikeuchi Nagoyadai meiyo kyōju' ['Drones are cruel weapons that have no regrets', warns Ikeuchi, Professor Emeritus at Nagoya University]

AS 2023 = 2023 02 24 'Ukuraina shinkō ichinen "senjō" hirogeru IT – gunmin no kyōkai o kuzusu ayausa' [One year after the invasion of Ukraine, IT is expanding the 'battlefield': danger of dissolving military-civilian boundaries]

Mainichi shinbun

MS 2013 = 2013 12 01 'Robotto no heitai' [Robot soldiers]

MS 2020 = 2020 08 11 'Bōkyaku no dorōn sensō' [The forgotten drone war]

MS 2022 = 2022 11 09 'Kōgekigata mujinki naze tayōsareru no? Anka de shippai sukunai, mikata ni higaisha nashi' [Attack drones: why are they used so much? Cheaper, fewer failures, no friendly fire]

MS 2023 = 2023 01 10 'Heiwa kokka' wa doko e? (7 shi) Minkaneiai gijutsu, dōnyū kasoku "shikikan handan o shien"' [Where is the peace state headed? (7: final) Civilian AI technology: accelerating introduction 'to assist commanders in their decision-making']

Yomiuri shinbun

YS 2001 = 2001 01 24 'Tetsuwan atomu no seiki: Kagaku to no taiwa: Yume no katachi (jō) Hinichijōteki seimei kōgaku' [The century of Tetsuwan Atomu: dialogue with science – the shape of dreams (I) unusual bioengineering]

YS 2015 = 2015 07 09 'Jinkō chinō to eiga' [Artificial intelligence and cinema] YS 2019a = 2019 06 02 'Satsujin robo kisei giron no ura' [Killer robots: behind the regulation debate]
YS 2019b = 2019 08 23 'AI heiki no kisei kōgeki handan o yudaneru beki denai' [Regulation of AI weapons: the decision to attack should not be left to them]
YS 2021 = 2021 06 06 '"AI heiki sensō no mirai" Ruisu A. Derumonte cho ['AI, Weapons and the Future of War', by Luis A. Del Monte]
YS 2022 = 2022 01 23 'AI heiki nobanashi ni dekinai Takushokudai kyōju Satō Heigo 55' [AI weapons cannot be left unchecked, says Professor Satō Heigo, 55, of Takushoku University]

The Reality of (Past) Future Air Combat Systems: Climate Wars, Carbon Costs and Rare Earth Elements

Jutta Weber

Introduction

The Intergovernmental Panel on Climate Change (IPCC) Working Group III warned in April 2023 that the small window of opportunity to meet the Paris Agreement's 1.5°C limit was rapidly closing. The effort to keep the 1.5°C target alive 'requires a significant step-change in an effort to phase out fossil fuels, build an electrified global energy system, end deforestation and tackle methane emissions' (IPCC, 2023).

What is often overlooked is that the world's militaries are responsible for at least 5.5 per cent of the total greenhouse gas (GHG) emissions (Parkinson and Cottrell, 2022; Peil, 2023a), which is more than the aviation and shipping industries combined (2 per cent each; Rajaeifar et al, 2022). A total of 87 per cent of the emissions come from the G20 nations. 'Residential emissions from military bases and emissions of naval operations are also significant. For arms-exporting or high military spending countries, such as the US and the UK, their indirect military emissions from the "value chain" can reach volumes similar to their direct annual emissions' (Michaelowa et al, 2022: iii). A report on the European Union (EU) military sector 'estimate[s] that the carbon footprint of EU military expenditure in 2019 was approximately 24.8 million tCO2e,[1] which is equivalent to the CO2 emissions of about 14 million average cars. We consider this a conservative estimate, given the many data quality issues' (Parkinson and Cottrell, 2021: 7). However, there is no reliable figure

[1] Ton of CO_2 equivalent.

for military carbon emissions because data is underreported or missing altogether (Belcher et al, 2019; Parkinson and Cottrell, 2021; Crawford, 2022). The military is often excluded from emissions reporting or only partially included in IPCC reports. In addition, most available calculations do not include the carbon costs of the destruction of natural or humanmade carbon stocks during warfare, reconstruction costs, healthcare for victims and so on.[2] One of many examples is the recent Israel-Gaza conflict which has caused immense human suffering, with more than 1,200 mostly Israeli civilians killed by Hamas as well as tens of thousands of Palestinians killed by the Israeli Defence Forces, 60–70 per cent of them civilians. More than half of the buildings in Gaza were destroyed (OCHA, 2024). This war is not only a humanitarian catastrophe but also has massive planet-warming carbon costs while climate impacts are already threatening Palestine's water and food security. A conservative study estimates the carbon emissions of the 60 days of the conflict at more than 281,000 metric tons of CO_2, equivalent to the annual emissions of 20 climate-vulnerable nations (Neimark et al, 2024).

Looking at another example, there are no targets for military GHG emissions in Germany or the EU that are consistent with the 1.5°C level, although the Paris Agreement was declared constitutionally binding by Germany's Federal Constitutional Court. At the same time, global military spending in 2023 was higher than in the previous three decades, with the largest increase being observed in Europe. It amounted to US$2,240 billion. This military spending is partly justified by the war in Ukraine. Conservative politicians and theorists alike point out that the geopolitical situation has changed fundamentally. China and Russia, in particular, are said to be rearming massively. According to the Stockholm International Peace Research Institute, the US spent US$916 billion on its military, accounting for 37 per cent of global spending, while China spent US$296 billion on its military, accounting for 12 per cent of the global budget. Even Russia, despite its ongoing war with Ukraine, accounted for 4.5 per cent of the global military budget in 2023 at US$130 billion. The combined budgets of China and Russia are less than half of the US budget alone. Much of the European spending is expected to be on so-called new weapons systems such as Future Combat or Global Combat Air Programmes. This increase

[2] For example, the burning of the oil wells alone during the Gulf War is estimated to make up 2–3 per cent of the total emissions in 1991 (Linden et al, 2004). But there are missing methodologies to calculate most of carbon as well as GHG emissions in general – for example, burning down a city or a forest as well as the costs for their rebuilding, or using chemicals such as napalm or uranium-depleted ammunition – though there are increasing efforts to develop these methodologies (Cottrell, 2022).

in global military budgets will most likely lead to a significant increase in GHG emissions.[3]

Against this background, I will discuss the carbon costs, GHG emissions and the rare earth metal dependencies of Future Combat Air Programmes and the realities of their future development and deployment. I suggest that the huge emissions and the dependencies will ultimately inhibit the realization of these systems, making their future something that has already passed.

1. The Future Combat Air System

The *Future Combat Air System* (FCAS) – Europe's largest defence project for the next few decades – is conceived as a network consisting of a 6th-generation jet fighter, remote carrier vehicles/uncrewed drones as well as sensors and weapons, all to be linked via an air combat cloud across all domains (Airbus, n.d.a; *Wikipedia*, n.d.; Bundeswehr, n.d.). This makes it similar to the US Next Generation Air Dominance (NGAD) system. It also promises to integrate and eventually replace current platforms such as Eurofighter or Rafale (Keisinger and Koch, n.d.). FCAS is being promoted as a 'Next Generation Weapon System' (NGWS) of uncrewed and autonomous platforms, and is being developed by Germany (Airbus), France (Dassault Aviation) and Spain (Indra Sistemas). Its development costs are estimated at €300–500 billion (Marischka, 2023). Another similar but smaller system is the BAE Tempest, which is being developed by the UK, Japan and Sweden. Both systems are expected to be operational by 2040, although some estimate that FCAS will not be available before 2050 (if at all). At the same time, it seems difficult to understand why the EU (and the UK and Japan) is supporting two competing projects that are extremely cost-intensive. Recently there have been rumours that Germany might abandon FCAS and join the BAE Tempest project (Moody, 2023). As it is unclear at the moment (at the time of writing in 2024) whether the projects will merge, I will focus on the Franco-German-Spanish FCAS project in this chapter, although most of my considerations and the problems discussed will apply for both (and many more) military development projects.

FCAS is promoted by Airbus as 'a key instrument in ensuring future European autonomy and sovereignty in defence and security. Furthermore,

[3] 'In response to the Russian invasion of Ukraine in February 2022, several European NATO member states had, by the end of March 2022, announced plans to raise military expenditure to reach or exceed the NATO spending goal of 2 per cent of gross domestic product (GDP) or more. These member states included Belgium, Denmark, Germany, Lithuania, the Netherlands, Norway, Poland and Romania. The acquisition of new weapon systems will probably be at the centre of these spending plans' (Tian et al, 2023).

it is said to strengthen Europe as an industry and technology hotspot, not just in the defence sector, but with important spillovers into the civilian world' (Airbus, n.d.a). Although it is very ambitious, it seems to lack a capability profile and is not based on combat scenarios, which is difficult for a system that is being developed now but will probably be deployed around 2040–2080. It seems to be driven mainly by technoimaginaries of military superiority and military dominance, an interest in technological leadership in general, the promotion of industry and exports, and competition between European nations, especially between Germany and France (see Peil, 2023a, 2023b).

2. Denying climate change: the carbon costs of the Future Combat Air System

One of the most significant GHG military emissions is from combat aircraft. For example, the US F-35 fighter is estimated to burn 5,600 litres of fuel per flight hour, which is 60 per cent more than its predecessor, the US F-16 fighter (Akkerman et al, 2022: 26). However, the development, construction and testing of these systems will also obviously produce a lot of GHG. Although there have been recent attempts to develop appropriate methodologies for accounting military emissions (Cottrell, 2022), there are as yet no calculations available to estimate the carbon costs of developing these complex systems over a course of 20 years. With deployment costs calculated at five times the development costs, the GHG emissions will even be tremendously higher *after* 2040/2050. Another major source of GHG in FCAS will be the development and deployment of AI applications in the combat cloud and drones: 'On the software side, building models for natural language processing and computer vision [as well as machine learning in general] is enormously energy hungry, and the competition to produce faster and more efficient models has driven computationally greedy methods that expand AI's carbon footprint' (Crawford, 2021: 15).

Although FCAS is expected to generate a tremendous amount of GHG emissions, this aspect is not addressed on the project's website, even though it is promoted as the first military technology development project to be accompanied by a 'working group on technology responsibility for FCAS' (Keisinger and Koch, n.d.). However, Airbus, one of the main developers of FCAS, nevertheless claims on its website: 'Our objective is to enable all our commercial and military aircraft and helicopters to be able to operate on 100 percent SAF [sustainable aviation fuel] by 2030' (Airbus, n.d.c; see also Airbus, n.d.b). Similarly, Rolls Royce as one of the partners for the development of the BAE Tempest system promises to 'revolutionise power for Combat Air Systems' via developing synthetic fuels or electrification (Rolls Royce, 2022).

While these new combat systems are far *more* energy-intensive than their predecessors, these promises of climate-neutral aviation are at the very least highly speculative if not greenwashing rhetoric: replacing fossil fuels with SAF can only reduce the greenhouse effect by a third (Clausen, 2022). The indirect greenhouse effect of aviation is permanent, regardless of whether synthetic fuel or hydrogen is used, because burning fossil or synthetic fuel at high altitude produces a significant greenhouse effect that is three times higher than fuel consumption on the ground:

> The idea that aviation will continue to be powered by combustion engines at high altitudes is questionable from a climate science perspective. This is because only one-third of the problem is fuel production; two-thirds of the problem lies in the fact that combustion emissions at high altitude are far more harmful in the stratosphere than on the ground. The same applies to water in the stratosphere as an 'exhaust gas' from the hydrogen fuel cell. Even water emissions at high altitudes would contribute significantly to the greenhouse effect. (Clausen, 2022: 5, translated by Joey Museba)

What most aviation companies deny – whether they operate in the civilian or military sector or both – is that a sustainable, climate-friendly technology policy should not only avoid the combustion of aviation fuels and develop *real* options for climate-neutral flying or reduce flying altogether.

Given the extreme energy consumption of fighter jets, electrification does not seem to be a real option for the coming decades. In 2022, the US Air Force (USAF) carried out the first test flight with an electric aircraft with a range of about 460 km and a top speed of 280 km per hour. An F-35 fighter has a range of 'more than 1,350 miles on internal fuel (1,200+ nautical miles), unlimited with aerial refueling' (Air Force, 2014) with a maximum speed of 1.6 Mach (~1,900 km per hour). Given these parameters, it is more than unlikely that FCAS will be an electrified system in the coming decades.

The airline industry has nevertheless repeatedly claimed that it will soon be carbon-neutral, but it also admits that its concepts for reducing GHG are essentially science fiction:

> Imagine a future where our customers are able to produce their own synthetic fuel supply (literally using carbon captured from the air, and hydrogen generated from water). Our customers will become resilient to fuel market supply shocks, and drastically reduce their logistics' footprint … We are actively pioneering technologies to make this Sci-Fi concept a reality. (Rolls Royce, 2023)

Many scientists see carbon capture and underground storage as risky and 'not proven at scale' (Anderson et al, 2023). The massive environmental problem of hydrogen at high altitude has already been discussed earlier.

There is obviously a lack of satisfactory documentation on the GHG emissions of the military in general and its existing deployed systems, but there are also no estimates of the GHG emissions of the much-touted future air combat systems: how much will the development and deployment of these systems produce? What is the overall impact of deploying these systems in the early 2040s when global warming is estimated to be around 2.0°C. Given the literally breathtaking kerosene consumption of jet fighters and the fact that there are no realistic alternatives to burning fossil fuels now or in the near future, it seems that the projections of future air combat systems largely ignore the issue of climate change: perhaps precisely because there are no solutions to the problem. But silencing these issues is as much a political move as the strategic assumptions, national politics and the like that go with it.

3. How the arms race will undermine the ecological transition: rare earth metals and extractivist politics

FCAS and similar systems rely heavily not only on fuels but also on rare earth metals (REE), which are at the same time necessary for the ecological transition (for example, for the production of batteries, wind turbines or solar cells). REE also have huge environmental and social costs: 'Mining is where we see the extractive politics of AI at their most literal. The tech sector's demand for rare earth minerals, oil, and coal is vast, but the true costs of this extraction is never borne by the industry itself' (Crawford, 2021: 15). Many scientists have pointed to the high social and environmental costs of extractive operations (Chagnon et al, 2022). These range from land, soil and water degradation (if not contamination), depletion of raw materials and natural resources to climate change, species extinction, biodiversity loss and deforestation. The costs are often borne by Indigenous or local populations whose quality of life is diminished or who are even forcibly displaced in the affected areas while the profits flow mainly to the high-tech companies. The extraction and manufacture of REE involves significant amounts of toxic and radioactive materials. One of the most dramatic examples is a humanmade lake near the world's largest REE mine, Bayan Obo in Inner Mongolia, which is also considered to be one of the most polluted and thereby dangerous places in the world. The negative environmental impacts of REE mining activities and processing can also be observed in the US, Brazil, India and Malaysia (Balaram, 2019). Therefore, it seems important to look for alternative sources and in the case of REE production to ensure that the ecosystem is protected (Drobniak and Mastalerz, 2022).

Rare earths are geopolitically contested. As the electrification and digitalization of our world accelerates, so too does the demand for REE used in these processes, such as the specialized rare earth magnets used in smartphones, wind turbines and photovoltaics, but also in drones, fighter jets and military robots. In 2023, China, the biggest producer of REE, restricted exports of resources such as gallium and germanium, which are used in the production of computer chips (while chip exports to China have recently been restricted by the US). Now restrictions on rare earths are also feared. New rare earths mines have been discovered in Sweden and there are other huge fields known to exist in Greenland, the US, Russia and Brazil, but China processes at least 85 per cent of the world's rare earth ore.

Among other things, carbon-fibre composites are needed for the flaps, tail and wings, the airframe, the nozzle as well as the postcombustion and propulsion of a jet fighter. Nickel is used for the flaps, the airframe and the engine. The electro-optical systems require aluminium, beryllium, cadmium, cobalt, copper, gallium, germanium, hafnium, indium, mercury, neodymium, tantalum, tellurium, titanium dioxide and yttrium (Peil, 2023a). To build a US F-35 jet fighter, 417 kg of rare earths are required (Selwyn, 2020: 26) and it is 'composed of 300,000 individual parts and assembled from 1,900 suppliers around the globe' (Selwyn, 2020: 35).

According to a 2020 study by the European Commission:

> the EU defence industry relies on the use of a wide range of materials with unique properties that make them essential for the manufacture of components used in military applications because the use of substitutes does not always guarantee the same performance. For example, REEs are indispensable in remotely piloted aircraft systems, precision guide munitions, targeting lasers and satellite communications. Rare earths are produced almost exclusively in China, which raises concerns not only on potential supply disruptions but also on strategic security. (Bobba et al, 2020: 69)

It is clear that end users outside China are dependent on China's global REE value chain (mining, oxides, metals, alloys and magnets), while the 'rising annual global demand for neodymium and dysprosium will significantly exceed annual global production by 2030' (Peil, 2023a). Although the amount of nickel and cobalt needed for energy storage in batteries may soon be reduced, there is nevertheless an increasing need for high purity nickel (for example, for cathodes) in arms production, which will lead to an even tighter nickel market. The same applies to cobalt which is not scarce per se, but is largely controlled by the markets of the Democratic Republic of Congo and especially China.

Apart from the problems of dependence on China's REE production, it is also likely that the massive increase in military spending, especially by the G20 countries, will lead to a global shortage on the REE market. This could intensify geopolitical crises and also lead to a massive shortage of supply on civilian markets, as the arms industry may be prioritized in the procurement of the necessary raw materials.

The ecological transition in the civilian world could be slowed down by the shortage of REE in civilian markets as the demand for REE and other resources increases in a thriving, expanding arms industry. This will hamper efforts to reduce GHG emissions in civilian production, which in turn will massively increase the socioeconomic costs of a (further) delayed transition: more natural disasters, more deaths, more biodiversity loss, more climate refugees and so on. Given the shortage of REE, the arms industry will 'acquire raw materials of specific qualities at a higher price sector than the competitive civilian market can afford. And this in turn means that civilian innovations, e.g., in the fields of renewable energies, e-mobility and ICT, either cannot be realised at the planned (and ecologically urgently needed) speed or, if alternative materials are used, not with the planned effectiveness' (Peil, 2023a).

The circular economy is seen as a possible solution to the shortage of REE and other materials in batteries. 'This includes themes such as *design for disassembly* – ensuring we have a better ability to access, recycle and repurpose materials through life, and ultimately reduce our exposure to future volatile supply chain risks' (Rolls Royce, 2023). However, according to a study by the European Commission, REE recycling is still 'in the early stages of development and face inherent difficulties: many other devices contain less than one gram of valuable REEs; the product design is unfriendly and not suitable for the easy separation of components, which makes the recycling process expensive. In addition, there is insufficient information on the REE content of different products' (Bobba et al, 2020: 33). Efficient and sustainable recycling processes do not yet exist – these 'processing techniques are yet to be established to exploit REEs via recycling' (Dushyantha et al, 2020: 1).

Conclusion: Westlessness, technological futures, new power weapons and geopolitics

Every day the news reports on record heatwaves, extreme droughts, severe storms, major floods, crop failures and other disasters caused by climate change on our planet. Few people still doubt the climate emergency, and it is becoming more deadly for all species every year. The amount of carbon dioxide in the atmosphere is already higher than it has been in the last three million years. It is estimated that there will be one billion climate migrants by 2030 (Vince, 2022).

Given these facts, it is hard to understand that European and other G20 countries are investing more and more in the arms race without considering the ecological consequences. Future combat air systems are an important part of the research and development effort that has huge carbon and GHG costs, and drive the race for rare earths, metals and other resources that are urgently needed for the ecological transition to keep our planet liveable. An ecological transition of aviation – military or civil – is nothing but an illusion: the announcement of clean hydrogen to the possibility of electrified fighter jets in the near future are totally unrealistic technoimaginaries – if not fairy tales. These future combat air systems are an irresponsible waste of scarce resources.

But after Russia's recent military aggression against Ukraine, many people seem to believe that today's accelerated arms race is inevitable. The argument is always that superpowers like China or Russia are investing heavily in armaments. But, as I mentioned earlier, the US alone spends 39 per cent of the world's military budget, and North Atlantic Treaty Organization (NATO) members spent a total of US$1,232 trillion in 2022. Russia invested US$86 billion and China US$916 billion (Tian et al, 2023). Even if a certain percentage of NATO's spending is invested in supporting Ukraine, it is still necessary to explain why NATO has such a huge (and growing) budget.

Today's weapon systems – especially nuclear ones – are far more destructive than those of the Cold War. Many disarmament treaties have not been renewed recently. In February 2019, for example, the US terminated the 30-year-old Intermediate-Range Nuclear Forces (INF) Treaty. In general, it seems that superpowers such as the US, the EU and Russia are increasingly less engaged in diplomatic efforts. This development is driven not least by the imaginary of 'Westlessness' in US and EU security discourse – a leitmotif of the declining hegemony of the 'West':

> Since 2013, security discourse has constructed a so-called Westlessness … As the Munich Security Report 2020 outlines, 'far-reaching power shifts in the world and rapid technological change contribute to a sense of anxiety and restlessness. The world is becoming less Western … Especially after Trump's 2016 presidential election in the United States … The US appeared to detach from alliances and dependency relations, "shrill warnings of an American Empire … have given way to fears that Uncle Sam might disengage from the world" (Munich Security Report, 2015: 6)'. (Ruppert, 2023: 8)

EU strategic security and policy papers called for Europe (and Germany) to become a new global power and to strengthen Europe's military power and weapons technology (Ruppert, 2023: 9).

Another reason may be the techno imaginaries of unbeatable AI weapons – from swarms of drones and autonomous weapon systems to data-driven warfare and total omniscience. The technopolitical imaginaries of military superiority, information dominance and omniscience are back again, which will not enhance our global security and peace. On the contrary, new technologies such as autonomous weapon systems and other AI-based systems have led to a shortening of response times, fuelling mutual insecurities and mistrust. Faced with the technoimaginary of unbeatable AI weapons, I would expect global actors to engage more in diplomatic negotiations and new arms treaties instead of constantly increasing military spending – especially at a time of global climate crisis. But this is clearly not the case.

It is an old insight of feminist peace as well as science and technology studies (STS) that the best way to peace is that of a global domestic policy that focuses not on armament, but on water and food security, education, a well-functioning health system and solutions for the climate crisis. It needs a caring foreign policy to resolve conflicts, not an arms race (Ruppert, 2022).

However, given the close links between the military, warfare and global climate problems, I doubt that FCAS and similar systems will ever be completed, but their development and testing will be fuelling the climate crisis. The latter may be a 'threat multiplier' (Crawford, 2022: 203), but when we run out of resources not only in terms of fossil fuels and REE but also water, food and other resources, we might reconsider the inevitability of highly sophisticated future combat systems and the trillions of dollars spent for the arms race in a world that is already in desperate need of reconstruction and not only a high-tech but also a socioecological transition to cope with the accelerating climate crisis.

Acknowledgements

I would like to thank Jascha Bareis, Jens Hälterlein and Thomas Christian Bächle for their helpful comments on earlier versions of this chapter. I would also like to thank Joey Museba for his proofreading and his translating some paragraphs of the German articles into English. Last but not least, I would like to thank Merle Spellerberg (the Greens) and Ralf Stegner (Social Democratic Party of Germany [SPD)] for the parliamentary breakfast on autonomous weapon systems on 25 May 2023 where I outlined some of my thoughts. And also thanks to Member of the European Parliament (MEP) Gesine Lötzsch for her comments which encouraged me to continue along this line.

References

Airbus (n.d.a) 'Future Combat Air System (FCAS): Shaping the future of air power'. Available from: https://www.airbus.com/en/products-services/defence/multi-domain-superiority/future-combat-air-system-fcas [Accessed 2 January 2025].

Airbus (n.d.b) 'Decarbonisation: towards low-carbon air travel for future generations'. Available from https://www.airbus.com/en/sustainability/respecting-the-planet/decarbonisation [Accessed 2 January 2025].

Airbus (n.d.c) 'Sustainable aviation fuel: advancing the ecosystem for a proven alternative fuel'. Available from: https://www.airbus.com/en/sustainability/respecting-the-planet/decarbonisation/sustainable-aviation-fuel [Accessed 4 December 2024].

Air Force (2014) 'F-35A Lightning II fact sheet'. Available from: https://www.af.mil/About-Us/Fact-Sheets/Display/Article/478441/f-35a-lightning-ii/ [Accessed 4 December 2024].

Akkerman, M., Burton, D., Buxton, N., Lin, H.-C., Al-Kashef, M. and de Vries, W. (2022) *Climate Collateral: How Military Spending Accelerates Climate Breakdown*. Transnational Institute (TNI) report. Available from: www.tni.org/en/publication/climate-collateral [Accessed 4 December 2024].

Anderson, K., Buck, H.J., Fuhr, L., Geden, O., Peters, G.P. and Tamme, E. (2023) 'Controversies of carbon dioxide removal', *Nature Reviews Earth and Environment*, 4: 808–814. https://doi.org/10.1038/s43017-023-00493-y

Balaram, V. (2019) 'Rare earth elements: a review of applications, occurrence, exploration, analysis, recycling, and environmental impact', *Geoscience Frontiers*, 10(4): 1285–1303.

Belcher, O., Bigger, P., Neimark, B. and Kennelly, C. (2019) 'Hidden carbon costs of the "everywhere war": logistics, geopolitical ecology, and the carbon boot-print of the US military', *Transactions of the Institute of British Geographers*, 45(1): 65–80. https://doi.org/10.1111/tran.12319

Bobba, S. et al (2020) *Critical Raw Materials for Strategic Technologies and Sectors in the EU: A Foresight Study*. European Commission, Joint Research Centre. Available from: https://data.europa.eu/doi/10.2873/58081 [Accessed 4 December 2024].

Bundeswehr. (n.d.) 'FCAS – Future Combat Air System'. Available from https://www.bundeswehr.de/de/organisation/luftwaffe/aktuelles/fcas-future-combat-air-system [Accessed 4 December 2024].

Chagnon, C.W. et al (2022) 'From extractivism to global extractivism: the evolution of an organizing concept', *Journal of Peasant Studies*, 49(4): 760–792. https://doi.org/10.1080/03066150.2022.2069015

Clausen, J. (2022) *Das Wasserstoffdilemma: Verfügbarkeit, Bedarfe und Mythen*. Borderstep Institut. Available from: https://www.borderstep.de/publikation/clausen-j-2022-das-wasserstoffdilemma-verfuegbarkeit-bedarfe-und-mythen-berlin-borderstep-institut/ [Accessed 4 December 2024].

Cottrell, L. (2022) *A Framework for Military Greenhouse Gas Emissions Reporting. Military Emissions Gap – June 2022*. Conflict and Environment Observatory. Available from: https://ceobs.org/wp-content/uploads/2024/04/CEOBS_A_framework_for_military_GHG_emissions_reporting.pdf [Accessed 4 December 2024].

Crawford, K. (2021) *Atlas of AI Power: Politics, and the Planetary Costs of Artificial Intelligence*, New Haven: Yale University Press.

Crawford, N.C. (2022) *The Pentagon, Climate Change, and War: Charting the Rise and Fall of U.S. Military Emissions*, Cambridge, MA: MIT Press.

Drobniak, A., and Mastalerz, M. (2022) 'Rare earth elements: a brief overview', *Indiana Journal of Earth Sciences*, 4. https://doi.org/10.14434/ijes.v4i1.33628

Dushyantha, N. et al (2020) 'The story of rare earth elements (REEs): occurrences, global distribution, genesis, geology, mineralogy and global production', *Ore Geology Reviews*, 122. https://doi.org/10.1016/j.oregeorev.2020.103521

IPCC (Intergovernmental Panel on Climate Change) (2023) *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.* IPCC, Core Writing Team, H. Lee and J. Romero (eds). https://doi.org/10.59327/IPCC/AR6-9789291691647

Keisinger, F., and Koch, W. (n.d.) 'What is the Future Combat Air System?', *FCAS Forum*. Available from: https://www.fcas-forum.eu/en/mission/ [Accessed 4 December 2024].

Linden, O., Jerneloev, A. and Egerup, J. (2004) *The Environmental Impacts of the Gulf War 1991.* IIASA Interim Report, International Institute for Applied Systems Analysis. Available from: https://pure.iiasa.ac.at/id/eprint/7427/ [Accessed 4 December 2024].

Marischka, C. (2023) *FCAS. Ansatzpunkte für eine Kampagne am Beispiel Stuttgart.* Informationsstelle Militarisierung e.V. (IMI). Available from: https://www.imi-online.de/2023/01/25/fcas/ [Accessed 4 December 2024].

Michaelowa, A. et al (2022) *Military and Conflict-Related Emissions: Kyoto to Glasgow and Beyond*. Zurich Open Repository and Archiv (ZORA) report. https://doi.org/10.5167/uzh-230045

Moody, O. (2023) 'Germany may abandon €100bn fighter jet project with France', *The Times*, 1 November. Available from: https://www.thetimes.co.uk/article/germany-may-abandon-100bn-fighter-jet-project-with-france-2n7ffv5g5 [Accessed 4 December 2024].

Munich Security Report (2015) *Collapsing Order, Reluctant Guardians?*]. Munich Security Conference report. Available from: https://securityconference.org/publikationen/munich-security-report-2015/ [Accessed 4 December 2024].

Neimark, B., Bigger, P., Otu-Larbi, F., and Larbi, R. (2024) 'A multitemporal snapshot of greenhouse gas emissions from the Israel-Gaza conflict'. Social Science Research Network (SSRN). Available from: https://ssrn.com/abstract=4684768 [Accessed 4 December 2024].

OCHA (United Nations Office for the Coordination of Humanitarian Affairs) (2024) 'Hostilities in the Gaza Strip and Israel – reported impact | Day 60. United Nations'. Available from: https://www.ochaopt.org/content/hostilities-gaza-strip-and-israel-reported-impact-day-60 [Accessed 4 December 2024].

Parkinson, S., and Cottrell, L. (2021) *Under the Radar: Europe's military sectors dodge scrutiny under European Green Deal*. The Left in the European Parliament, Conflict and Environment Observatory, Scientists for Global Responsibility. Available from: https://left.eu/issues/publications/under-the-radar-europes-military-sectors-dodge-scrutiny-under-european-green-deal/ [Accessed 4 December 2024].

Parkinson, S., and Cottrell, L. (2022) *Estimating the Military's Global Greenhouse Gas Emissions*. Conflict and Environment Observatory report. Available from: https://ceobs.org/estimating-the-militarys-global-greenhouse-gas-emissions/ [Accessed 4 December 2024].

Peil, K.H. (2023a) 'Environmental Impact of €100 billion in arms investments: Problem outline and short study', Informationsdienst Umwelt und Militär. Available from: https://umwelt-militaer.org/bws-en/ [Accessed 4 December 2024].

Peil, K.H. (2023b) *Umweltauswirkungen von 100 Milliarden Euro für Rüstungsinvestitionen: Problemaufriss und Kurzstudie*. Rosa Luxemburg Stiftung, Informationsdienst Umwelt und Militär. Available from: https://info.umwelt-militaer.org/pdf/2023/2023-02_BwS_Web.pdf [Accessed 4 December 2024].

Rajaeifar, M.A. et al (2022) 'Decarbonize the military – mandate emissions reporting', *Nature*, 611(7934): 29–32. https://doi.org/10.1038/d41586-022-03444-7

Rolls Royce (2022) 'How we are revolutionising power for Combat Air Systems'. Available from: https://www.rolls-royce.com/media/our-stories/discover/2022/tempest_electrical.aspx [Accessed 4 December 2024].

Rolls Royce (2023) 'Building a sustainable future for global defence operations'. Available from: https://www.rolls-royce.com/media/our-stories/discover/2023/building-a-sustainable-future-for-global-defence-operations.aspx [Accessed 4 December 2024].

Ruppert, L. (2023) 'Geopolitics of technological futures: warfare technologies and future battlefields in German security debates', *Geopolitics*, 29(2): 581–606. https://doi.org/10.1080/14650045.2023.2174431

Ruppert, U. (2022) 'Ein bisschen Feminismus im Krieg? Über das Paradoxe feministischer Außenpolitik', *PROKLA. Zeitschrift für kritische Sozialwissenschaft*, 52(208): 499–507. https://doi.org/10.32387/prokla.v52i208.2010

Selwyn, D. (2020) *Martial Mining*. London Mining Network (LMN). Available from: https://londonminingnetwork.org/project/martial-mining-2020/ [Accessed 4 December 2024].

Tian, N. et al (2023) *Trends in World Military Expenditure, 2023*. Stockholm International Peace Research Institute (SIPRI). https://doi.org/10.55163/CQGC9685

Vince, G. (2022) *Nomad Century: How to Survive the Climate Upheaval*, London: Allen Lane.

Wikipedia (n.d.) 'Future Combat Air System'. Available from: https://en.wikipedia.org/wiki/Future_Combat_Air_System [Accessed 4 December 2024].

Showcasing Power, Performing Responsibility? Introducing Military Artificial Intelligence Discourses in China

Thomas Christian Bächle and Xiran Liu

Introduction

Major geopolitical crises of the recent past and present underline both a changing global security architecture and its strong dependence on and connectedness to military technologies. Shifts in the global balance of power, the vulnerabilities of democratic processes, transforming media systems or the rise of populism – these tendencies also have strong repercussions for the impact of *information warfare*, the targeted use of propaganda to influence public opinion or spin the narratives on real events. Besides the strategic use of communication, often more prominently discussed connections between warfare technologies and artificial intelligence (AI) concern the *automation* of target selection, engagement and decision making in *kinetic* or *cyber* (for example, attacks on digital infrastructures) domains.

Against this background, this chapter will first give a short introduction to the broader framework of warfare technologies and their relationship with broader AI narratives. Autonomous weapon systems (AWS) and military AI can be understood as a subset of larger AI discourses that share similar functions (such as techno-solutionism), as well as a strong connection to political communication aimed at both national and international audiences and without (such as deterrence or the construction of national identity). This reiterates that current discourses and debates on AI do not simply account for particular technical advancements in automation, data analysis or machine learning. They also exceed the reflections on the social and political ramifications these developments might have. Hereby, they also go beyond the many regulatory, legal or ethical questions that are commonly associated with them. AI discourses reflect imagined versions of the future, often recounted along the lines of doom and gloom, techno utopian and dystopian visions of destruction. In astute awareness of these discourses, AI is utilized as a signifier by particular actors (state, corporations, interest groups and so on) to communicate meanings in order to produce particular effects in public discourses, policy making or research and development (section 1).

In a second step, the chapter zooms in on military and security-related applications of AI in the Chinese context. Military- and security-related AI is to be understood as a subset of narratives in a stylized 'AI race' between global powers. This also marks AI as a highly politicized technology, which is related to questions of national identity and is utilized as a tool in political communication. Military AI and AWS embody a specific imagination, tied to narratives of AI supremacy, and also interwoven with ideas of clean 'digital warfare', more rational, more precise, more just even – increasing efficiency through automation while reducing the cost of human casualties. AWS epitomize this idea of a clean warfare, and yet shed a light on the stark disconnect between the shiny imaginations of an AI-enabled warfare and the reality of the brutal and destructive industrial warfare that causes the unimaginable human suffering in current crises.

The case of China with regard to cutting-edge military technology is particularly interesting, as it offers another perspective to the theme of geopolitical dominance, which is often primarily seen through the prism of Euro-American interpretations. The same is true of the potential functions these representations and imaginaries might have in political communication, both domestically and in the global geopolitical arena, which includes the relationship to national identity. When looking at a particular set of imaginaries, it must be noted that the reality of the political system in China is characterized by a lack of accountability: it is not necessary for policy makers or government officials in China to tell their citizens what they are doing and to what ends they are doing it. There is no direct need to accurately inform the public or seek democratic legitimacy for the development, deployment or regulation of technologies.

Still, any attempt to portray China as a villainous geopolitical antagonist would only serve to permeate Euro-American discourses. A mere critique of People's Republic of China (PRC) policies would hardly surprise anyone, which is why this chapter, in a third step, provides a small but nuanced introduction to major media discourses on military AI. It will look at some of the patterns of media representations and showcase how they tie in with larger themes of national identity, policies, international relations or technological risks and promises. While other nations can certainly be analyzed in a similar fashion, what makes the case of China special is that the narratives surrounding technology are highly politicized and actively utilized to inform the public through images that shine a positive light on the central government. Another factor that makes the case of China so worthwhile for the discussion of narratives and imaginations is the relative impermeability in (at least) a linguistic (as opposed to the global use of English, for example) and a technological sense. The latter particularly includes the state-controlled media system, which is even more dominant in the shaping of public discourses as cross-border internet traffic is limited and the Chinese market offers its own digital platforms (section 3).

1. AI in Euro-American and Chinese contexts: conceptual debates and their performative qualities

It is a common understanding that interpretations and representations of technology vary historically and culturally. Technologies that are associated with the notion of AI are no exception in this regard. The AI-related discussions in Euro-American contexts are often representative of value-based approaches to technologies, which are at the same time meant to affirm particular images of national identity.

The capitalist rationale of AI technologies are countered with critical conceptualizations that reflect on the negative effects of universal datafication and new forms of power imbalances (Mejias and Couldry, 2019), surveillance (Zuboff, 2019), political discourses, public opinion or democratic processes. Closely tied to these imaginaries are common normative value discourses on AI in Euro-American contexts, such as fairness, accountability and transparency (FAT), algorithmic bias and value-based design (Simon, Wong and Rieder, 2020) or more generally the debates on 'AI Ethics' (Coeckelbergh, 2020), 'AI for Social Good' or 'Trustworthy AI' (Bareis, 2024).

AI in many instances assumes the role of a signifier that takes particular meanings – often in an instrumental sense – which includes its status as an object, its 'thingness' (Suchman, 2023). Besides the more abstract and value-based framings that are associated with AI, the imaginaries on digital technologies also function as reassurance of a political or cultural identity, conveyed and shaped by particular technologies and their meanings. European national identities or a pan-European identity, for example, are created and maintained with particular reference to a larger geopolitical order, namely European economic and political competition with the US and China (Monsees and Lambach, 2022). In this narrative of competition, China – portrayed as a systemic rival – dwarfs the US in terms of (value-based) data protection, infrastructural dominance or espionage. The latest political tropes of this are 'derisking' and 'strategic decoupling'.

AI narratives and imaginaries in China are both similar to and different from those found in Euro-American contexts. Similarities between Euro-American and Chinese discourses largely stem from a globally shared vocabulary in discussing AI, the meanings of which, however, sometimes differ starkly depending on their cultural context. What adds to this complex dynamic is that while AI narratives and imaginaries differ globally (Cave and Dihal, 2023), the technological developments and research and development goals in most cases come with a distinct political component: their own 'performative politics' (Bareis and Katzenbach, 2021). In other words, the meanings associated with these technologies are actively shaped, often as a means to an end. The vocal debates on AI in policy making, academia, the industry and civil society all contribute to the typically normative and value-based discussions. They do not just characterize the technology itself but also fulfil a performative function, to reassure the idea of a national identity, to engineer a sociopolitical vision of governance or to perform symbolic dominance.

Besides the similarities, as part of the attempt to reflect on what can be seen as distinct in Chinese discourses on technology, there is a list of explanations that serves as a common frame of reference. While the reference points of distinction serve an epistemological purpose, they run the risk of reiterating simplistic narrative structures, such as West vs. East, evoking themes of techno-orientalism or a clash of civilizations that is ideological in nature but AI-driven this time around (McInerney, 2024). One of the most striking patterns that is often employed in both public discourses and research refers to the different features of collectivistic (the Chinese model) and individualistic (the European and North-American model) societies. While the former prioritizes the common good of the community, the latter gives credence to personal freedoms and development. Another somewhat simplisitic pattern that follows a positive/negative evaluation in public discourse and also in research neatly ties AI either to the idea of progress and economic growth (good AI) or to surveillance and social control practices (bad AI).

Against these patterns, the Confucianism-framed discourses on harmony and community in China are somewhat complemented with the sociopolitical flipside of forced conformity and the absence of the Euro-American idea of individuality and personal freedoms (Sterckx, 2019).[1] Analyses of US and Chinese AI policies substantiate – and in doing so reiterate – these schematic explanations, comparing a 'Protestant ethic', which emphasizes 'the individual's responsibility and rights', to 'Confucianism', which 'trusts leaders to guide the people towards a harmonious society' (Hine and Floridi,

[1] Regarding the individual/collective divide, a literal reading of Confucian teachings suggests understanding the notion of harmony as more related to the family unit rather than the flourishing of the nation (see, for example, Ebrey, 2009: 17–26). Nonetheless, the tropes of harmony and peaceful coexistence are heavily referenced as guiding China's international relations approaches, as seen in the idea of building 'a community with a shared future for humankind' (Ministry of Foreign Affairs, 2024).

2024: 270). What the US and China have in common is that both nations share the same driving force (namely being a 'world leader in AI') and goal of hegemonic influence – what is particular for the Chinese context is that innovation is also seen as posing a risk to social stability (Hine and Floridi, 2024: 270).

As in Europe and North America, AI is commonly seen as a tool and driver to transform traditional industries in the Chinese context as well. In academic discourses the most discussed issues about AI ethics also reflect familiar territory from a Euro-American perspective, as they feature ideas such as privacy, equality, security, freedom, autonomy, unemployment, legality and transparency (Zhu, 2024: 1265). The shared ethical concerns point to similar AI-related problems, including privacy intrusion, algorithm discrimination or the challenge of allocating responsibility (Zhu, 2024: 1271).

This is reflected in AI policy and regulation frameworks that become increasingly complex and also mirror the attempts of Western legislation to implement ethical principles in the development of new technologies. European Union (EU), US and Chinese AI policies all 'share the idea that AI must be used for the good of humankind, and that it must be used in ways that are safe, transparent, equitable, and responsible' (Qiao-Franco and Zhu, 2022: 16). This is also true for the conceptual level, which evokes the seemingly universal ideas of fairness, justice or responsibility and extends to the somewhat to be expected principles of harmony and friendliness. However, major differences are still palpable in the prioritization of values, which see 'responsibility over freedom, obligation over rights, and the group over the individual' (Gal, 2020; Qiao-Franco and Zhu, 2022: 9). Chinese policy making may be quick to codify these ethical principles into regulations on AI applications, such as the laws on generative AI introduced in 2023 (Zhang, 2023). However, the de jure deliberations (and their performative quality – see the discussion later on in the chapter) are not necessarily reflective of the de facto practises that for the most part are not backed by civil society and lack effective oversight (see also Qiao-Franco and Zhu, 2022).

This disconnect clearly speaks to the performative dimension of regulatory debates, which are often intended to portray and communicate particular meanings in order to frame policy-making processes in the public arena – both nationally and internationally – and in doing so shape the imaginary dimension of the technology. The objective is not only to engineer legislation or steer research and development in a desired direction, but also to perform particular meanings within the national or global realms of political communication. The active meaning making can range from stimulating heightened acceptance of technologies to deterrence or social control by dominating public discourses. Therefore, AI is also part of what in political theory is referred to as 'securitization':[2] the construction of security-related issues in this performative dimension via speech acts. When issues are constructed as threats, they are moved from policy- to security-related discourses, shifting the power dynamic in governing and policy making.

In China, this is reflected in (at least) two corresponding tendencies in the acts of framing and utilizing AI, namely security politics (relating to securitization) and governance practices. First, according to Zeng (2022: 4), 'AI is being securitized by the Chinese central government to mobilise local states, market actors, intellectuals and the general public'. The 'historical anxieties about technology and regime security' are identified as constituting factors of the 'security discourse in China's AI politics', which is reinforced by geopolitical competition.[3] Second, regarding governance, AI is heavily integrated into the central government's aspirations to execute power via digital technologies, which includes exerting social control over its citizens through universal surveillance and the often-referenced social credit system (Xu et al, 2022). The Chinese efforts to establish AI-enabled forms of 'authoritarian governance' are threefold: pushing for an AI-driven and internationally competitive digital *economy*; a technologically enabled *social transformation* in the age of AI and digital means; and also utilizing AI *rhetorically* in efforts to be 'proving ideological superiority of its authoritarian and communist values' (Zeng, 2022: 5). The presentation of these efforts to promote AI technologies can also be assumed to promote and foster the authoritarian government's self-proclaimed legitimacy in the eyes of its citizens. This is particularly important in a system that chooses and assesses its officials on meritocratic principles. As in other areas, the legitimacy of the political order is infused by citizens' perception of (good) governance performance (Dong and Kübler, 2021), a form of output legitimacy (Schmidt, 2020). This includes not only official policies but also the management of images and narratives on current events (for the case of COVID-19, see Klenk and Gurol, 2024).

[2] In 'Securitisation Theory', the 'Copenhagen School argues that security is not a given but it is constructed through inter-subjective social and discursive interactions between powerful actors who propose definitions of threats and relevant audiences who acknowledge these definitions' (Stępka, 2022: 18).

[3] Despite common Euro-American interpretations of Chinese politics, Zeng (2022: 4) states that 'China's AI approach is sophisticated and multifaceted' and its 'AI plans are primarily driven by contestation and the struggle for resources among domestic stakeholders who are economically motivated and have little awareness of the bigger geopolitical picture'. Zeng contends (2022: 4) that contrary to existing analysis, 'China's AI approach is not a top-down geopolitically driven nationally concerted strategy'. However, the current geopolitical realities make it more likely that the top-down approach will be increasingly enforced. Presumably AI will play an ever-increasing role not just as governance technology but also as a rhetorical device.

These efforts are reflected in the major policies that shape China's AI strategy. The 'New Generation Artificial Intelligence Development Plan' (AIDP, 2017), published by the State Council in 2017, 'acts as a unified document that outlines China's AI policy Objectives' (Roberts et al, 2021: 60) and 'highlights three areas where AI can make a substantial difference within China: international competition, economic development, and social governance' (Roberts et al, 2021: 62). The economic development and social governance parts discuss the use of AI as a solution in areas such as social prosperity, coping with an ageing population or environmental issues. Part of social engineering is the social credit system, which is intended to regulate the behaviour of its citizens. The government's efforts to actively shape AI discourses is also reflected in the ways in which AI is framed in state media with the goal to shape its public perception. Despite these efforts, Cui and Wu (2021: 54) emphasize that still, 'as a controversial technology, the public image of AI is inevitably associated with risks'.

Besides the discourses found in mass media, social media can be seen to play a double role: they can (or are envisioned to) be platforms where policy debates occur and undermine the central government's official stance – but at the same time are being utilized by state actors to steer public opinion via censorship or propaganda (DeLisle, Goldstein and Yang, 2016; Mao and Shi-Kupfer, 2020). Zeng et al (2020: 332) comment that while 'an over-hyped and economy-focused coverage of AI' is characteristic of both Euro-American and Chinese media, the difference in China lies in 'the continuing absence of vocal and influential communities that reveal the "blind spots" within the current AI discourse'. At the same time, despite the role that is sometimes attributed to them, Zeng et al (2020: 332) found that 'social media's role as a counter-public sphere in AI discourses is minimal'. Other scholarship still emphasizes that there are 'vibrant discussions' on Chinese social media platforms, with diverse voices, including those by 'scholars, IT industry actors, journalists, and members of the general public', which are argued to be 'a valuable source for understanding the future trajectory of AI development in China as well as implications for global dialogue on AI governance' (Mao and Shi-Kupfer, 2023: 373). More recent studies of the public perception of AI – effectively demonstrated on a global level by ChatGPT (see, for example, Lian et al, 2024) – showed an overall critical stance towards AI technologies, especially disinformation risks, unemployment and the human/computer relationship.

Overall, there seems to be a complex dynamic between the necessity felt by an authoritarian system to legitimize technological change and the efforts to unify sociotechnical imaginaries in a media system that potentially also facilitates counternarratives. This picture becomes even more complex when technologies are expressly relevant for security issues or discourses are actively utilized in conflicts.

2. Military AI in Chinese policy making and public discourses

The Russian aggression against Ukraine that started in 2022, an increasingly volatile situation in the Middle East, and the expansive gestures by China towards Taiwan and in the South China Sea – a long list of hot and cold conflicts can also be read as symptoms of larger geopolitical shifts. Slowly succeeding a bipolar order, the past two decades have not only seen the rise of new players that are striving for power in the global arena, most notably the BRICS nations (Brazil, Russia, India, China and South Africa). The model of the Western-style capitalist democracy has also been subjected to internal crises with the rise of populism and democratic backsliding (Müller, 2016; Scheppele, 2018).

These geopolitical shifts coincide with technological developments that are utilized in conflicts, ranging from deterrence to actual acts of aggression, seemingly widening the range of what is commonly referred to as hybrid warfare. First, this concerns an array of *functions, domains and weapons capabilities that seems to go beyond the means of industrialised warfare*. Attacks are launched in the cyber-realm, with the goal to disrupt and sabotage critical infrastructures and services (Arquilla, 2021; Perlroth, 2021). Machine-learning algorithms, which analyse patterns in large datasets, are combined with probability-based automated decision making (commonly referred to as 'AI') that has also made a critical impact in military contexts, concerning both the automated selection of targets and automated attacks in the kinetic realm (Scharre, 2018). Second, this becomes relevant as *information warfare*, which includes attempts to dominate a discourse in order to achieve a particular effect within the public sphere (Seib, 2021; Dutton, 2023). This discourse can concern the shaping of public discourse and war propaganda, but it also refers to metanarratives such as 'clean' warfare through AI.[4] Technology is also utilized besides its actual functionality for deterrence when framed as a means to outsmart rivals and enemies. Showcasing technological advancements directed to an international audience is one of the strategies that is used for this purpose. A robot, resembling a dog with an automatic rifle mounted on its back, used during the China-Cambodia 'Golden Dragon 2024' military exercise is one of the more recent examples, which made headlines in European and American media. Unlike algorithm- and data-based automated decision-making systems, military robots provide the opportunity to perform and demonstrate the impact AI can have on military power by employing visual

[4] This can be understood as a continuation of the promise of the virtualization of warfare, see Bareis and Bächle, 'The Realities of Autonomous Weapons' in this volume.

means. The materiality of AI emphatically references known objects to make the message heard, a type of military skeuomorphism.

These large-scale tendencies translate into many fields of policy making in China. In 2003 the Chinese Communist Party (CCP) and the People's Liberation Army (PLA) introduced the 'Three Warfares' doctrine, including strategies to shape public opinion through media, to apply psychological means to influence an opponent or legal warfare that deliberately seeks to find loopholes in regulatory frameworks or international conventions to expands one's sphere of influence (Jackson, 2015). The endeavour to effectively regulate AWS is an example of how AI is a major point of reference not just when it comes to questions of functionality, but also in these larger discourses. Human control or machine autonomy is ambivalent in Chinese contexts (Qiao-Franco and Bode, 2023). One of the major sites where different meanings of AI-enabled military technologies are performed and enacted is the Convention on Certain Conventional Weapons (see CCW, 2017; CCW, 2018; Sayler and Moodie, 2020), an international regulatory forum of the United Nations (UN) that seeks to regulate autonomous weapons. While the different nations are trying to find an agreed-upon definition of what AI in the military domain is and subsequently how it should be regulated, there is always the performative dimension (cf. earlier discussion in this chapter): it becomes visible in the gap between the practicalities of defining the technologies in question and taking concrete steps of regulation on the one hand, and the imaginations of AI technologies in policies and public statements on the other (see Bächle and Bareis [2022] for a detailed discussion).

China's long-term military objectives include 'achieving world-leading defence industrial status by the end of the 2040s' (Janes, 2022: 3). A 2022 report on China's military capabilities assesses that the 'Chinese defence industrial base has made considerable strides over the past two decades towards the objectives outlined in the defence White Papers to develop advanced military capabilities (which, according to the White Papers, is intended to enable the Chinese military to win high-intensity, information-centric local wars)'. But it also emphasizes that 'the defence industrial base is still short of being able to develop and produce high-quality core technology components and systems that are available elsewhere in the world (notably in the US and Europe) that is the yardstick with which Beijing continues to measure capability' (Janes, 2022: 5).

Regarding AI and cyber-capabilities in particular, the AIDP – while outlining global corporate competition in more general terms – also explicitly refers to AI-related effects and promises associated with weapons technology: 'Rather than outspending the US in conventional weaponry, China considers investing in AI as an opportunity to make radical breakthroughs in military technologies and thus overtake the US' (Roberts et al, 2021: 62). These so-called ‘asymmetric tactics’ include cyberwarfare, with China developing focussed capabilities, while at the same time promoting ‘international initiatives for regulating hostile state-run activities in cyberspace, and to fill the existing regulatory gap for state behaviour in this domain’ (Roberts et al, 2021: 63).

In the information warfare domain, China is also intensifying its efforts to utilize media technologies to steer public discourse towards the fulfilment of its interests. A report by the Atlantic Council[5] underlines the belief in the Chinese political elites ‘that Western countries, and especially the United States, have been able to exert global dominance because they possess what China terms “discourse power” (话语权): a type of narrative agenda-setting ability focused on reshaping global governance, values, and norms to legitimize and facilitate the expression of state power’ (Thibaut, 2022: 2). One of the most salient measures is the idea of ‘cyber sovereignty (网络主权)’, which ‘in China’s definition’ refers to ‘the right of each country to exert total control over the Internet within its borders’ (Thibaut, 2022: 3). In propaganda messages, deliberately targeted to international audiences, ‘China often targets audiences with narratives that erode the legitimacy of the liberal democratic framework and that resonate with local experience’ (Thibaut, 2022: 3). It is a deliberate attempt to curtail what is regarded as Western norms and values, such as privacy, free speech or freedom of information – while applying ‘the very mechanisms the US and its allies created to govern and shape a ‘free, open, secure, and interoperable’ digital world’ (Thibaut, 2022: 3). Besides social media activities (for example, using bots to amplify official narratives), the Central Propaganda Department (CPD) is key to achieving this goal, as it ‘is responsible for regulating the content of China’s publishing, news media, and film industries, and for providing content directives for Chinese state media organizations like People’s Daily’ (Thibaut, 2022: 13).

Against this background, the following analysis will focus in particular on the ways in which military uses of AI are represented in the Chinese media. It will present major tendencies in *People’s Daily Online*[6] as a major state-controlled medium. The overarching patterns include current representations of military AI (section 3.1), shifting meanings in recent years in light of geopolitical developments (section 3.2) and how these

[5] The Atlantic Council is a self-described non-partisan US-based think tank in the field of international affairs.

[6] *People’s Daily* can be dubbed the ‘mouthpiece of the Central Committee of the Communist Party of China (CPC), its editorials and commentaries represent the official viewpoints of the Chinese authorities’ (Zeng et al, 2020: 320) and is hence particularly relevant when analysing the portrayal of military technology.

representations relate to national (section 3.3) or international (section 3.4) aspects of communication.

3. Military AI discourses: core themes in the Chinese media

Amid global discussions surrounding AI across various sectors, the topic of AI in the military – particularly regarding autonomous weapons – does not seem to have gained significant traction on Chinese internet platforms, either in state-sponsored official outlets or on social media. This relative lack of discussion, especially when contrasted with the more widespread dialogue around AI or military as isolated topics in general, is evident in the sparse search results on major media platforms, such as *People's Daily Online* (people.cn / 人民网).

People.cn, the online portal of the *People's Daily*, is the official newspaper of the Chinese Communist Party (CCP). The online portal features original *People's Daily* articles, while also frequently selecting and republishing pieces from other state-owned media, such as the *Beijing Daily, People's Liberation Army Daily*, *China Military*, *Xinhua Net* and *Guangming Daily*. Therefore, *People's Daily Online* can be positioned as a key source for content that reflects the central government's official stance. Despite the 2017 launch of the AIDP, which highlights AI development as a national priority, searches on *People's Daily Online* for terms like 'military AI' and 'autonomous weapons' return fewer than 20 articles annually.

Similarly, on *Zhihu*,[7] a popular Chinese social question-and-answer platform with over 220 million users, military AI topics see little engagement. For instance, the top threads associated with 'military AI' such as the question 'What are the applications of AI in the military?'[8] have received only 18 responses, and the answers tend to be short and not sophisticatedly structured, providing minimal material for in-depth analysis. Meanwhile, another question – 'What attitudes should individuals and states take toward lethal autonomous weapons?' – has garnered just one answer. Such low engagements may reflect a general lack of public interest in extensively

7 *Zhihu* (知乎, literally meaning 'do you know?') functions similarly to the US site *Quora*, where the engagements centre around user-initiated questions for others to provide answers. *Zhihu* stands out in the Chinese social scene for its role as a knowledge market, where users often engage in extensive and sophisticated discussions on complex topics. This makes *Zhihu*'s user profiles and popular topics different from other types of social media platforms, such as microblogging (for example, *Weibo*), where engagements tend to be shorter and more spontaneous.

8 Contents are written in simplified Chinese; the translations from Chinese to English are provided by the authors.

engaging with military AI-related topics on *Zhihu*, despite other popular topics on the platform typically drawing thousands of responses and active discussions.

This review aims to contextualize China's AI policies and practices within the state-controlled media environment. To this end, around 50 news articles relevant to military AI from China's state media outlets were identified and collected using a list of search terms.[9] The articles, published between 2017 and 2024, encompass a range of formats, including coverage of events, opinion pieces and commentary. As a preliminary effort to map some notable trends in terms of how the state media constructs the concept of military AI, a qualitative approach was chosen to review the collected articles.

The analysis identifies four key themes and trends regarding how military AI is discussed by state media outlets, which are explored further later on with references to specific articles: a cautious official stance on endorsing military AI and a clear non-endorsement of autonomous weapons (section 3.1); an increasingly cautious approach from 2018 through 2024 (section 3.2); the strategic use of this caution to build China's image as a responsible global power (section 3.3); and positioning China as an objective observer through a distancing from discussing military AI development and application in and by China, often linking the subject to specific foreign countries (section 3.4).

3.1 Current stances towards military AI expressed through state media

State media coverage of military AI and autonomous weapons consistently and continuously reflects an official stance of caution towards the concept of military AI, which is manifested in avoiding the term 'military AI' when describing domestic issues and in refraining from portraying the application of AI in the military in a positive light; in addition, a non-endorsement of the development and application of autonomous weapons is evident.

This cautious stance is explicitly conveyed through opinion and commentary pieces that elaborate on the political, technical and ethical dangers associated with the overuse of military AI, as well as the catastrophic consequences of deploying lethal autonomous weapons. For example, a commentary published by the *People's Liberation Army Daily* on 12 September

[9] The list of search items used in the collection of data from *People's Daily Online* (people.cn / 人民网) and other state media include 'military AI' ('人工智能军事'), 'autonomous weapon' ('自主武器'), and 'unmanned system' ('无人系统'). The list was curated based on terminologies used in official documents from international governing bodies such as the UN (for example, CCW, 2016; Gill, 2018) or the ICRC (for example, ICRC, 2017). See also the subsequent discussion in this chapter of linguistic idiosyncrasies of military AI.

2024, titled 'Artificial intelligence accelerates the process of weapon autonomisation',[10] highlights the existing extensive use of autonomous weapons on the battlefield. It argues that these AI-powered systems challenge geopolitical stability by disrupting the current balance of military power (Pei et al, 2024). The article stresses the ethical risks posed by autonomous weapons that can independently select and engage targets without human intervention, effectively delegating the power to determine life and death to machines (Pei et al, 2024). Such applications could lead to increasing desensitization to the cruelty of warfare, resulting in a gamification of conflict and potential overuse of military force, thereby significantly challenging established ethical standards (Pei et al, 2024). Just like with the broader ethical issues raised with AI (see the earlier discussion on this), these questions and concerns are also shared in the Euro-American debates on autonomous weapons, mirroring the commonly evoked US, International Committee of the Red Cross (ICRC) and UN talk of 'select and engage targets without human intervention' and also being reflected in notions such as 'meaningful human control'.

In addition to explicit commentary labelling military AI as a catalyst for significant danger and, thereby, expressing disapproval of the application of autonomous weapons, the official stance of caution and non-endorsement is also implicitly conveyed through news reports on international events that question the legitimacy of such technologies and advocate for stringent regulatory measures. While these reports may appear to merely present discussions from these events, the frequency and depth of coverage suggest that the central government supports these critical perspectives or at least subscribes to the performance of responsibility and regulatory theatre that it also maintains on international platforms (cf. earlier discussion on this).

A notable example of such event reporting in an international arena is the coverage of the 2024 Vienna Conference on Autonomous Weapons Systems held in April 2024. Multiple prominent national media outlets reported on the event, including *China Central Television* (CCTV), *Xinhua Net* and *Beijing Daily*. During the conference, Austria's Foreign Minister, Alexander Schallenberg, likened the current state of AI in military development to the 'Oppenheimer moment' of our generation (Guo and Liu, 2024), referencing J. Robert Oppenheimer's creation of the atomic bomb during the Second World War to highlight the danger of autonomous weapons. Schallenberg warned that while technology is advancing rapidly, politics is

[10] The news articles from China's state media outlets referenced in this chapter are written in simplified Chinese. All translations of news article headlines and content from Chinese to English in this chapter are provided by the authors.

failing to keep pace with it, and autonomous weapons are poised to dominate future battlefields.

This reference to the 'Oppenheimer moment' was repeatedly quoted in China's national media coverage, often appearing in headlines to emphasize the severe risks associated with military AI and autonomous weapons. For instance, an article by *China Military*, an official news outlet sponsored by the Chinese People's Liberation Army, entitled 'Are AI weapons facing the "Oppenheimer moment"?' cited 'relevant comment' ('相关评论'),[11] which argued that deploying AI-driven lethal AWS in combat could spark a 'third military revolution' akin to the introduction of gunpowder and nuclear weapons (Guo and Liu, 2024). The article notes that controlling these weapons in the short term is challenging as global conflicts escalate, while countries and companies increasingly promote and invest in AI, complicating regulatory efforts (Guo and Liu, 2024).

The article concludes with a seemingly neutral remark, stating that 'to make good use of AI weapons as "double-edged swords", broad international cooperation and a deep understanding of future responsibilities are needed to ensure their safety and control' (Guo and Liu, 2024, para. 8). However, the choice to highlight the challenges of regulating ethical AI use in the military underscores the official stance of caution regarding the application of AI in a military context that the Chinese government generally promotes on international platforms.

This overall expressed caution and non-endorsement can be interpreted as framing military AI as a threat by state media, aligning it with 'securitization' efforts by the Chinese government (section 1). At the same time, it is at odds with the efforts of information warfare conducted on multiple layers (section 2). In this case, framing military AI as a threat shapes public opinion and political discourse to prioritize security concerns. Such framing therefore has the potential to shift power to those responsible for managing security and national defence, reinforcing a top-down, state-centric approach to AI development.

3.2 Shifting portrayals of military AI over time

Searches of China's major state media outlets reveal that news articles specifically addressing military AI began to emerge in 2017. Following the release of the AIDP, coverage increased in 2018, interestingly coinciding with the onset of the US-China trade war. However, from 2019 to 2022,

[11] Citing a 'relevant comment' is a common journalistic practice in state media, where statements or comments are used without clear attribution (for example, expert opinions, discussion points at a conference) to support previously made claims in an article.

the coverage on military AI declined, before surging again in 2023 and 2024. A review of these articles shows that, despite maintaining a consistent cautious and non-endorsement stance, the tone of caution has subtly but noticeably intensified over the years.

On 8 November 2018, the *Chinese Liberation Army Daily* acknowledged the significance of AI in military applications in an article entitled 'The military application of AI is a double-edged sword'. It begins by emphasizing how traditional military powers, such as the US and Russia, recognize the immense potential of AI in military contexts, viewing it as a 'game-changing' disruptive technology (Xiong, 2018). The article proceeds to discuss the development and application of AI in the military with caution, describing the integration of AI into military affairs as a 'fierce' ('来势汹汹') trend (Xiong, 2018). It highlights the ethical dangers posed by the future deployment of numerous intelligent unmanned systems on the battlefield, and subscribes to narratives of how AI may significantly reduce the cost of warfare and make the prospect of 'zero casualties' for combat personnel a reality. This situation could lead military powers to use force more arbitrarily (Xiong, 2018). From a technical perspective, the article warns that in complex battlefield environments, highly intelligent unmanned combat systems may face challenges such as misidentification, communication degradation or even 'turning against' their operators when subjected to enemy electromagnetic or cyberattacks (Xiong, 2018). Furthermore, issues such as indiscriminate killing and system malfunctions raise ongoing concerns about the military applications of intelligent weapons (Xiong, 2018).

It is noteworthy that the comments made in this piece are general, avoiding mention of specific technologies or applications, such as unmanned aerial vehicles. As one of the early state media articles on military AI, it establishes a cautious approach to the technology's application in military contexts. Nevertheless, it also presents a relatively balanced view, acknowledging AI's potential to enhance weapon performance with the goal of 'effectively preserving national security' (Xiong, 2018). This perspective aligns with the metaphor of the 'double-edged sword' used in the headline.

The dominant tone in military AI-themed news articles published in 2018 reflects a cautious approach that also acknowledges the significant potential of AI applications in the military. This blend of scepticism and recognition may come as a surprise, especially given that China had just released the AIDP a year earlier, which positioned AI, including military applications, as a strategic priority. Despite this, state media did not engage with military AI as proactively as the plan might suggest. This contrast indicates that state media are conservative in addressing military AI in an overly positive light without further explicit endorsement from higher authorities.

A notable example that supports this pattern is a news article published by the *People's Liberation Army Daily* on 30 November 2018, following

China's President Xi Jinping's participation in the ninth collective study session of the Political Bureau of Chinese Communist Party (CCP) Central Committee.[12] During this session, Xi stressed the importance of encouraging scientists to boldly explore the uncharted frontiers of AI technology, ensuring that China leads in AI theoretical research and secures the strategic high ground in critical core technologies (Zhan, 2018). Citing Xi's statement, the article, entitled 'AI technology: the future will surpass our imagination', contrasts sharply with other pieces published that year to explicitly prioritize technological advancement in AI despite the associated risks, asserting that 'excessive worry can hinder the development of AI. Not long ago, the internet revolutionised most traditional industries, forcing everyone to either embrace the change or be left behind' (Zhan, 2018: para. 21). Such an optimistic tone from state-backed media outlets regarding the application of AI in a military context remains extremely scarce to date.

Compared to the relatively balanced stance towards military AI and occasional bold endorsements of such applications in 2018, concerns and caution seem to dominate the discussion of military AI in state media by 2023 and 2024. This shift is exemplified by an article published in the *People's Liberation Army Daily* on 6 September 2023 entitled 'The dilemma of AI weaponization'. While the title references AI weapons in general, the article primarily focuses on a subset of this technology – AWS. It begins by recounting a case presented at a UK conference on future air combat, where an AI-powered drone killed its operator to achieve a higher score (Zhang and Wang, 2023). The article then elaborates on how AI's algorithmic nature makes its military application a 'double-edged sword', raising concerns about the reliability of AI weapons and the potential for technical malfunctions with severe consequences (Zhang and Wang, 2023). The piece offers a more detailed and explicit caution against AI in military contexts, particularly expressing strong reservations about autonomous weapons (Zhang and Wang, 2023). While it also employs the 'double-edged sword' metaphor, similar to the previously referenced 2018 article, the sentiment in 2023 is much more focused on the dangers posed by military AI. Similarly, an article published on 13 August 2023 by *Guangming Daily* entitled 'AI weapons frequently appear on the battlefield, its development is concerning' highlights the alarming features of autonomous weapons. It cites the ICRC to emphasize the potential harm to civilians, noting that

[12] In China's current political system, President Xi Jinping is also the General Secretary of the Chinese Communist Party (CCP) and Chairman of the Central Military Commission (CMC). These collective study sessions cover a range of selected topics, and one key objective of these meetings is to 'send out signals about the current policy focuses and intentions of the central leadership' (Lu, 2007: 2). Therefore, Xi's statements during these sessions serve as core messages to members of the CCP, the sole ruling party in China.

such weapons can operate without human control over targets, posing a significant challenge to international humanitarian regulations (*Guangming Daily*, 2023). The article stresses the absence of an international treaty to prohibit AI weapons, despite the existence of regulations for nuclear and chemical weapons (*Guangming Daily*, 2023).

This type of coverage – primarily focused on the severe risks of military AI and autonomous weapons, alongside references to their use in real combat settings and heated international debates – seems to dominate state media articles on military AI in 2023 and 2024. On the one hand, the development of more concrete military AI technologies and applications over the years has made the associated risks and harms more apparent, potentially fuelling discussions and increasing articles focused on the downside of military AI. On the other hand, this heightened coverage coincides with a new wave in the global AI race, sparked by ChatGPT's attention-grabbing launch in late 2022 and China's reported pressure to catch up (Gordon, 2023). This may have prompted China to increasingly frame military AI as a threat to position itself as a responsible actor in the global AI landscape. While such positioning echoes the growing internal demand for development in AI ethics, it may also serve as an alternative narrative to advance its position in the global AI race (cf. earlier discussion on this). This topic will be explored in greater detail in the following section.

3.3 Military AI-related representation as part of national image

Given China's openly declared goal of achieving superiority in the global AI race by 2030 (Koetse, 2024), it may seem intuitive to assume that this ambition extends to advancements in military AI. One might also expect the Chinese state media to proactively report on China's development efforts in military AI to project a stronger national image. However, as indicated in the previous section, a review of existing media content reveals little evidence of China actively promoting its technical advancements in military AI. In fact, caution regarding such applications and a non-endorsement stance on autonomous weapons have become increasingly pronounced over the years, as demonstrated by state media coverage.

This situation contrasts with China's established patterns of enhancing its national image by showcasing military power in non-AI-related military areas, such as tactical training, advancements in fighter jets and missile technology through regular national coverage. Notably, there appears to be great caution in using the term 'military AI'. For instance, during the 70th anniversary celebration of National Day in 2019, which featured a major parade showcasing cutting-edge military advancements, the focus was predominantly on traditional military weapons rather than on AI technologies. Such parades, held every ten years, are considered the most important events for demonstrating the nation's military capabilities. During the parade, which showcased hundreds of pieces of military equipment, including deep-sea drones and intercontinental ballistic missiles, unmanned systems were highlighted for their roles in the battlefield. Specifically, unmanned aircraft were noted for applications such as target positioning, artillery calibration and damage assessment. Although it was acknowledged that unmanned and intelligent systems represent the future of warfare, the coverage exclusively used the term 'unmanned' and did not mention 'artificial intelligence' (for example, *people.cn*, 2019). Despite the rising importance of AI in the military and as a deterrent force, the performative theatre of the military parade puts a strong focus on the industrialised kinetic aspects of weapons that showcase machinistic force with visual magnitude.

Notably, the parade coverage repeatedly asserted that all showcased equipment had been independently developed by China (*people.cn*, 2019). This reflects a common approach in the Chinese context to highlight the country's ability to develop weapons and systems independently, with a strong emphasis on self-reliance. Such an approach points to a semantic difference that should be recognized to avoid misinterpreting messages conveyed through these media outlets.

In state media articles written in Chinese, it is common to see the words 'weapon' (武器) and 'autonomous/autonomy/autonomously' (自主) positioned closely together in sentences that promote Chinese military advancements. However, the official media's focus when using both terms has been consistently on stressing China's efforts and accomplishments in 'researching and developing *weapons autonomously*' ('自主研发武器') instead of 'researching and developing *autonomous weapons*'('研发自主武器'). In this context of 'researching and developing weapons autonomously', the term '自主' (directly translated as 'autonomously') could be more accurately understood as signifying independence and self-reliance in the weapon development process rather than the creation of weapons capable of self-assessment, targeting and combat without human intervention. This distinction in how 'weapon' and 'autonomous' are structured within sentences is important, as different structures convey distinct agendas.

While the parade coverage signifies a distancing from endorsing the concept of 'military AI', this approach can also be seen as a strategic effort to construct a positive national image: by taking a stance against provocative military AI development and application, China positions itself as a responsible superpower occupying a moral high ground in the global discourse on regulating this technology. This is evident in reports of China's participation in global initiatives aimed at regulating military AI, such as its involvement in the second summit on Responsible AI in the Military Domain (REAIM) held in Seoul, South Korea in September 2024. An article published by *Xinhua Net* on 11 September 2024 entitled 'Chinese representative: will work with the international community to reach more consensus on the military applications of artificial intelligence' highlights China's commitment to the safety and governance of AI development as an emerging AI superpower. Citing the Chinese representative at the summit, the coverage emphasizes China's call for collective action to 'resist all possible risks and challenges' posed by AI and its military applications (Feng, 2024: para. 4) .

The construction of China as a responsible player in the military AI discourse is also reflected in state media's framing of proactive military AI advancements as promoting irresponsible geopolitical agendas. A prominent example is a recent article entitled 'AUKUS accelerates military application of AI', published by *China Military* on 19 August 2024. AUKUS is a trilateral security partnership involving Australia, the UK and the US. The news article reports on a recent AUKUS trial featuring AI-powered drones and other autonomous combat devices, dedicating a section to characterizing AUKUS as a 'war-mongering' alliance (Li, 2024). It contends that while AUKUS claims its military AI advancements aim to address major-power competition, its true intention is to jeopardize regional peace and stability (Li, 2024). Additionally, the article emphasizes that AUKUS operates as a tightly bound alliance under US leadership, expecting its members to possess high-level capabilities across multiple domains and to align with US interests against major rivals (Li, 2024). This dynamic, the article explicitly warns, will exacerbate confrontations, increase the risk of weapon proliferation, and threaten peace and stability in the Asia-Pacific region (Li, 2024).

This framing of military AI advancements as indicative of ill-intentioned alliances underscores calls for stricter regulations on military AI, positioning supporters of such regulations such as China as morally and ethically superior. By portraying alliances like AUKUS as self-interested, the article also implicitly contrasts their priorities – often focused on individual rights and benefits, as seen in Western countries – with China's emphasis on social responsibility and community, which is characterised by a focus on harmony that underscores Chinese AI ethics (cf. earlier discussion on this).

3.4 Self-positioning in military AI discussions and references to other countries

The narratives adopted by state media often position China as an objective observer of international discussions and relevant military AI practices adopted by specific countries. This distancing aligns with the previously identified trend of refraining from endorsing the concept, as media coverage provides only minimal acknowledgement of China's advancements in applying AI to weaponry. For instance, the most explicit references to military AI in China are limited to its role in facilitating flight training, with no mention of combat capabilities. It would be arbitrary to attribute this minimal coverage solely to a deliberate distancing strategy, given other factors not related to propaganda such as confidentiality of weapon development details. Nevertheless, this lack of emphasis on military AI progress in China contrasts with the media highlights of technical advancements in other weapon systems, such as intercontinental missiles.

The narratives adopted by China's official outlets often reflect a disapproval of rival countries' provocative approaches to military AI. The US emerges as the most frequently mentioned country in news coverage by state-sponsored media outlets since 2018. A rough count shows that among news articles discussing military AI in 2023 and 2024 on *People's Daily Online*, at least one third deliberately focus on actions taken by the US military to advance AI applications, often adopting a highly critical tone. This somewhat expectedly reflects the main rivalry for global hegemonic dominance.

One strand of this criticism highlights the rapid advancement in AI weapon development by the US in technical terms. For example, an article entitled 'The US is establishing its overseas AI battle labs for the first time', published on 27 October 2023 by *China Military*, reports on the announcement by the US Department of Defense (DoD) to build two BRAVO AI battle labs in Europe and Indo-Pacific (US Department of Defense, 2023). The article criticises this move, labelling it an effort to 'stir up the AI military race' (Shi, 2023: para. 12). It further notes that various actions by the US and its allies, including the UK, Japan and South Korea, to advance AI in military contexts have raised concerns that centre around ethical issues, the safety of machine operations and the potential to escalate regional arms races (Shi, 2023).

Another line of criticism regarding the approach made by the US to advancing military AI centres on national policy support, which is regarded as provocative, particularly the loosening of regulations governing the development of military AI. A news article published on 10 February 2024 by *China Military*, entitled 'The US gives green light to developing artificial intelligence weapon systems', highlights that the DoD has amended its guidelines for AWS and this move is interpreted to facilitate their large-scale development and application. The article emphasizes that despite the DoD's explicit directive requiring commanders and operators of autonomous weapons to act with caution and adhere to the laws of war and relevant treaties, it does not specifically prohibit the development of any particular weapon systems (Nai, 2023). Furthermore, it does not impose restrictions on the possession of autonomous AI weapon systems; as long as these systems meet the design, testing and validation requirements outlined in the directive, they can be developed and deployed (Nai, 2023).

Both lines of criticism contribute to a broader narrative that portrays the US as an irresponsible nation prioritizing its own interests at the expense of global peace. The coverage consistently underscores how the US is expanding and overusing military AI overseas, suggesting unjustified interference in the sovereignty of other countries. This narrative is often implicitly constructed through articles discussing US military AI advancements, but it is also explicitly reflected in headlines such as ‘Has the Ukrainian battlefield become a testing ground for US military AI weapons?’, published on 6 May 2024 by *Beijing Daily*. Such narratives in the realm of military AI resonate with China’s repeated claims opposing interference in other countries’ internal affairs, particularly those actions taken by the US.

In contrast to such framing concerning China’s rival countries, coverage addressing military AI in South Africa presents a starkly different tone, reflecting the amicable relations between China and South Africa. An article entitled ‘South Africa establishes a national defence AI centre’, published by *China Military* on 31 May 2024, discusses the establishment of the Defence Artificial Intelligence Research Unit (DAIRU), South Africa’s fourth AI research centre and its first dedicated to defence. The article adopts a subtly encouraging tone, highlighting that South Africa’s AI technology is leading among African countries as one of the few nations on the continent with advanced AI infrastructures (Zhang, 2024). Notably, the article does not address the risks associated with the development and application of military AI, a topic typically elaborated upon in state media articles discussing international conversations or commenting on relevant practices by the US.

It is important to note that state media coverage of military AI focusing on countries other than the US is limited. While this single report on South Africa may serve only as anecdotal evidence, it offers a preliminary indication of the differing tones used to portray the adoption of military AI by rival versus allied countries.

Conclusion

AI debates and discourses in China are driven by many of the familiar vocabularies that are shared globally, including notions such as privacy, legality or transparency. Both in policy making and research these concepts are emphasized, while it remains doubtful to what degree they play an actual role in the de facto legal practices, considering the lack of democratic legitimacy or oversight. Therefore, the performative dimension of AI rhetorics gains major relevance: on the one hand, the authoritarian systems seeks legitimacy through an idealized notion of governance that is perceived as ‘good’ in a national context; on the other hand, on an international stage, AI tropes are employed to perform competitiveness and aspiration with reference to other nations.

The same performative mechanism applies to military AI. On a national level, the development of AI capabilities is connected to the promise of security, while on an international level, two general and contradictory tendencies emerge. AI is used to showcase military power to the effect of intimidation and deterrence, which is at odds with the official stance on military AI that is characterized by caution. The latter stresses the potential dangers and the need to regulate these new types of technologies to the effect of performing responsibility in an international arena of policy making and regulation.

The four themes identified through the analysis of how military AI and autonomous weapons are portrayed in state media reflect this official stance. On the one hand, state media increasingly frame the use of AI in the military as a potential threat rather than as a technically and ethically promising application by extensively discussing associated risks and concerns in their articles. This mirrors the same themes that can also be found in Euro-American discourses, which correspond with the general impression that military AI narratives and framings are shared across national discourses.

On the other hand, this framing is complemented by China's self-positioning as a responsible actor in the global AI landscape, with media coverage highlighting the country's participation in regulatory efforts and critiquing provocative advancements in military AI. The latter is exemplified by other nations (the US in particular) as the driver of a military AI race. This officially promoted image is at odds with many of the Chinese efforts to utilize AI technologies in competitive or conflict scenarios. They span from the development of actual AI-enabled warfare capabilities to utilising particular AI narratives and imaginations in shaping public realities in the interest of the Chinese government.

China's official stances in policy making and state media representations not only form a noteworthy contradiction with its actual military investments and activities, but such a contradiction may itself also be part of the larger effort of warfare manifested through psychological means. The deterring effects of military AI discourses thrive on insecurity, which is certainly not only true for China, but also other global powers such as the US, the EU and Russia. AI not only relates to an imagined future that comes with a disquieting vision of uncertainty in technological, ethical or sociopolitical terms; the question of what AI is potentially capable of blends with the risk scenarios of what the enemy is potentially capable of. Software, cyber, intelligence, kinetic and information – what follows from the almost universal functionality and applicability of AI in these realms of conflict, warfare and the attendant insecurity is that it can be creatively weaponized.

Acknowledgements

The authors wish to thank Jing Zeng, Jascha Bareis, Karoline Kozlowski and Christoph Ernst for their support and constructive comments on the first version of this text. In part, this chapter was written within the framework of the research network 'Meaningful Human Control – Autonomous Weapon

Systems between Regulation and Reflection', funded by the German Federal Ministry of Education and Research (grant number 01UG2206C).

References

AIDP (Artificial Intelligence Development Plan) (2017) *China's 'New Generation Artificial Intelligence Development Plan'*, G. Webster, R. Creemers, E. Kania and P. Triolo (trans.). Digichina, Stanford University. Available from: https://digichina.stanford.edu/work/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/ [Accessed 4 December 2024].

Arquilla, J. (2021) *Bitskrieg: The New Challenge of Cyberwarfare*, Cambridge: Polity Press.

Bächle, T.C., and Bareis, J. (2022) '"Autonomous weapons" as a geopolitical signifier in a national power play: analysing AI imaginaries in Chinese and US military policies',. *European Journal of Futures Research*, 10(20). https://doi.org/10.1186/s40309-022-00202-w

Bareis, J. (2024) 'The trustification of AI: disclosing the bridging pillars that tie trust and AI together'. *Big Data and Society*, 11(2). https://doi.org/10.1177/20539517241249430

Bareis, J., and Katzenbach, C. (2021) 'Talking AI into being: the narratives and imaginaries of national AI strategies and their performative politics', *Science, Technology, & Human Values*, 47(5): 855–881. https://doi.org/10.1177/01622439211030007

Beijing Daily (2024) '乌克兰战场已成美军的AI武器实验室？' [Has the Ukrainian battlefield become a testing ground for US military AI weapons?], 6 May. Available from: https://baijiahao.baidu.com/s?id=1798256245650808098&wfr=spider&for=pc [Accessed 4 December 2024].

Cave, S., and Dihal, K. (eds) (2023) *Imagining AI: How the World Sees Intelligent Machines*, Oxford: Oxford University Press.

CCW (Convention on Certain Conventional Weapons) (2016) 《禁止或限制使用某些可被认为具有过分伤害力或滥杀滥伤作用的常规武器公约》缔约方第五次审查会议 [*Fifth Review Conference of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects*] (CCW/CONF.V/10). Available from: https://digitallibrary.un.org/record/3856242 [Accessed 4 December 2024].

CCW (2017) *Report of the 2017 Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS)* (CCW/GGE.1/2017/CRP.1). Available from: https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Meeting_of_High_Contracting_Parties_(2017)/2017_CCW_GGE.1_2017_CRP.1_Advanced_%2Bcorrected.pdf [Accessed 4 December 2024].

CCW (2018) *Report of the 2018 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems* (CCW/GGE.1/2018/3). Available from: https://documents.un.org/doc/undoc/gen/g18/323/29/pdf/g1832329.pdf [Accessed 4 December 2024].

Coeckelbergh, M. (2020) *AI Ethics*, Cambridge, MA: MIT Press.

Cui, D., and Wu, F. (2021) 'The influence of media use on public perceptions of artificial intelligence in China: evidence from an online survey', *Information Development*, 37(1): 45–57. https://journals.sagepub.com/doi/full/10.1177/0266666919893411

DeLisle, J., Goldstein, A., and Yang, G. (eds) (2016) *The Internet, Social Media, and a Changing China*, Philadelphia: University of Pennsylvania Press.

Dong, L. and Kübler, D. (2021) 'Government performance, political trust, and citizen subjective well-being: evidence from rural China', *Global Public Policy and Governance*, 1: 383–400. https://doi.org/10.1007/s43508-021-00024-9

Dutton, W.H. (2023) *The Fifth Estate: The Power Shift of the Digital Age*, Oxford: Oxford University Press.

Ebrey, P.B. (2009) *Chinese Civilization: A Sourcebook*. New York: Free Press.

Feng, Y. (2024) '中国代表：将同国际社会就人工智能军事 应用问题形成更多共识' [Chinese representative: will work with the international community to reach more consensus on the military applications of artificial intelligence]. *Xinhua Net*, 11 September. Available from: http://world.people.com.cn/n1/2024/0911/c1002-40317684.html [Accessed 4 December 2024].

Gal, D. (2020) 'Perspectives and approaches in AI ethics: East Asia', in M.D. Dubber, F. Pasquale, and S. Das (eds) *The Oxford Handbook of Ethics of AI*, Oxford: Oxford University Press, pp 607–624.

Gill, A.S. (2018) '联合国对于应对致命自主武器系统领域新技术的作' [The role of the United Nations in addressing emerging technologies in the area of lethal autonomous weapon systems]. Available from: https://www.un.org/en/un-chronicle/role-united-nations-addressing-emerging-technologies-area-lethal-autonomous-weapons [Accessed 4 December 2024].

Gordon, N. (2023) 'ChatGPT's rise has China scrambling to catch up in an A.I. arms race that will determine global "wealth, power, and influence"', *Fortune*, 8 August. Available from: https://fortune.com/2023/08/08/chatgpt-china-catch-up-arms-race-global-wealth-power-influence/ [Accessed 4 December 2024].

Guangming Daily (2023) 'AI武器频现战场 发展前景引人忧' [AI weapons frequently appear on the battlefield, its development is concerning], 13 August. Available from: https://baijiahao.baidu.com/s?id=1774081112167885745&wfr=spider&for=pc [Accessed 4 December 2024].

Guo, B., and Liu, M. (2024, May 13) 'AI武器面临'奥本海默时刻'？' [Are AI weapons facing the 'Oppenheimer moment'?], *China Military*, 13 May. Available from: http://military.people.com.cn/n1/2024/0513/c1011-40234656.html [Accessed 4 December 2024].
Hine, E., and Floridi, L. (2024) 'Artificial intelligence with American values and Chinese characteristics: a comparative analysis of American and Chinese governmental AI policies', *AI & Society*, 39: 257–278. https://doi.org/10.1007/s00146-022-01499-8
ICRC (International Committee of the Red Cross) (2017) 自主武器系统：增强武器关键功能的自主性带来的影响 [*Autonomous Weapon Systems: Implications of Increasing Autonomy in the Critical Functions of Weapons*]. Available from: https://www.icrc.org/zh/publication/4283-autonomous-weapons-systems [Accessed 4 December 2024].
Jackson, L. (2015) 'Revisions of reality: the three warfares – China's new way of war', in *Beyond Propaganda. Information at War: From China's Three Warfares to NATO's Narratives*, London: Legatum Institute, pp 5–15. Available from: https://www.lse.ac.uk/iga/assets/documents/arena/archives/information-at-war-from-china-s-three-warfares-to-nato-s-narratives-pdf.pdf [Accessed 4 December 2024].
Janes (2022) 'China – defence industry', report in Janes *Jane's World Defence Industry 50*. Available from: https://shop.janes.com/products/directories/world-defence-industry-50-6541-3000220009 [Accessed 15 January 2025].
Klenk, E. and Gurol, J. (2024) 'The role of narratives for gaining domestic political legitimacy: China's image management during COVID-19', *Journal of Chinese Political Science*, 29: 337–359. https://doi.org/10.1007/s11366-023-09865-z
Koetse, M. (2024) 'In the race for AI supremacy, China and the US are travelling on entirely different tracks', *The Guardian*, 9 January. Available from: https://www.theguardian.com/world/2024/jan/09/in-the-race-for-ai-supremacy-china-and-the-us-are-travelling-on-entirely-different-tracks [Accessed 4 December 2024].
Li, H. (2024) "奥库斯'加快AI军事应用' ['AUKUS' accelerates military application of AI], *China Military*, 19 August. Available from: http://military.people.com.cn/n1/2024/0819/c1011-40301319.html [Accessed 4 December 2024].
Lian, Y., Tang, H., Xiang, M., and Dong, X. (2024) 'Public attitudes and sentiments toward ChatGPT in China: a text mining analysis based on social media', *Technology in Society*, 76: 102442. Available from: https://www.sciencedirect.com/science/article/abs/pii/S,,,0160791X23002476 [Accessed 4 December 2024].
Lu, Y. (2007). *The Collective Study Sessions of the Politburo: A Multipurpose Tool of China's Central Leadership* (Briefing Series, Issue 27). University of Nottingham. China Policy Institute. Available from: https://www.nottingham.ac.uk/iaps/documents/cpi/briefings/briefing-27-collective-study-sessions-of-the-politburo.pdf [Accessed 4 December 2024].

Mao, Y., and Shi-Kupfer, K. (2020) What's hyped and what's real: the state's innovation ambitions versus society's concerns – social media discussions on artificial intelligence (AI) and ethics in China', in M. Dowling et al (eds) *Deutsch-chinesische Innovationspartnerschaft: Rahmenbedingungen, Chancen und Herausforderungen. Die Policy Briefs der deutschen DCPI Expertengruppe 2017–2019*, Weimer: Metropolis Verlag, pp 377–390. Available from: https://hdl.handle.net/21.11116/0000-000A-94EC-5 [Accessed 4 December 2024].

Mao, Y., and Shi-Kupfer, K. (2023) 'Online public discourse on artificial intelligence and ethics in China: context, content, and implications', *AI & Society*, 38: 1–17. https://link.springer.com/article/10.1007/s00146-021-01309-7

McInerney, K. (2024) 'Yellow techno-peril: the "clash of civilizations" and anti-Chinese racial rhetoric in the US–China AI arms race', *Big Data & Society*, 11(2). https://doi.org/10.1177/20539517241227873

Mejias, U.A. and Couldry, N. (2019) 'Datafication', *Internet Policy Review*, 8(4). https://doi.org/10.14763/2019.4.1428

Ministry of Foreign Affairs. (2024). 'Wang Yi: building a community with a shared future for mankind has become a glorious banner leading the progress of the times', 7 March. Available from: https://www.mfa.gov.cn/mfa_eng/wjbzhd/202403/t20240308_11256416.html [Accessed 4 December 2024].

Monsees, L., and Lambach, D. (2022) 'Digital sovereignty, geopolitical imaginaries, and the reproduction of European identity', *European Security*, 31(3): 377–394.

Müller, J. W. (2016). *What Is Populism?*, Philadelphia: University of Pennsylvania Press.

Nai, Q. (2023) '美为发展人工智能武器系统'开绿灯" [The US gives green light to developing artificial intelligence weapon systems], *China Military*, 2 October. Available from: http://www.81.cn/gfbmap/content/2023-02/10/content_333354.htm [Accessed 4 December 2024].

Pei, S., Shi, H., and Huo, J. (2024) '人工智能加速武器自主化进程' [Artificial intelligence accelerates the process of weapon autonomization]. *People's Liberation Army Daily*, 12 September. Available from: http://www.81.cn/szb_223187/szbxq/index.html?paperName=jfjb&paperDate=2024-09-12&paperNumber=11&articleid=939489 [Accessed 4 December 2024].

People.cn. (2019) '32个装备方队动地而来 40%装备首次亮相' [32 equipment formations approach; 40% of the equipment showcased for the first time], 1 October. Available from: http://politics.people.com.cn/n1/2019/1001/c430388-31383122.html [Accessed 4 December 2024].

Perlroth, N. (2021) *This Is How They Tell Me the World Ends: The Cyberweapons Arms Race*, London: Bloomsbury.

Qiao-Franco, G., and Bode, I. (2023) 'Weaponised artificial intelligence and Chinese practices of human-machine interaction', *Chinese Journal of International Politics*, 16(1): 106–128, https://doi.org/10.1093/cjip/poac024

Qiao-Franco, G., and Zhu, R. (2022) 'China's artificial intelligence ethics: policy development in an emergent community of practice', *Journal of Contemporary China*, 33(146): 189–205. https://doi.org/10.1080/10670564.2022.2153016

Roberts, H., Cowls, J., Morley, J., Taddeo, M., Wang, V., and Floridi, L. (2021) 'The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation', *AI & Society*, 36: 59–77.

Sayler, K. M., and Moodie, M. (2020) *International Discussions Concerning Lethal Autonomous Weapon Systems* (Report IF11294). Congressional Research Service. Available from: https://fas.org/sgp/crs/weapons/IF11294.pdf [Accessed 4 December 2024].

Scharre, P. (2018) *Army of None: Autonomous Weapons and the Future of War*, New York: W.W. Norton and Company.

Scheppele, K.L. (2018) 'Autocratic legalism', *University of Chicago Law Review*, 85(2): 545–583. https://chicagounbound.uchicago.edu/uclrev/vol85/iss2/2

Schmidt, V.A. (2020) *Europe's Crisis of Legitimacy: Governing by Rules and Ruling by Numbers in the Eurozone*, Oxford: Oxford University Press.

Seib, P. (2021) *Information at War: Journalism, Disinformation, and Modern Warfare*, Cambridge: Polity Press.

Shi, W. (2023) '美在海外首建AI战斗实验室' [The US is establishing its overseas AI battle labs for the first time], *China Military*, 27 October. Available from: http://military.people.com.cn/n1/2023/1027/c1011-40104865.html [Accessed 4 December 2024].

Simon, J., Wong, P.-H., and Rieder, G. (2020) 'Algorithmic bias and the value sensitive design approach', *Internet Policy Review*, 9(4). https://doi.org/10.14763/2020.4.1534

Stępka, M. (2022) 'The Copenhagen School and beyond: a closer look at securitization theory', in *Identifying Security Logics in the EU Policy Discourse*, Dordrecht: Springer, pp 17–31.

Sterckx, R. (2019) *Chinese Thought: From Confucius to Cook Ding*, London: Random House.

Suchman, L. (2023) 'The uncontroversial "thingness" of AI', *Big Data & Society*, 10(2). https://doi.org/10.1177/20539517231206794

Thibaut, K. (2022) *Chinese Discourse Power: Ambitions and Reality in the Digital Domain*. Atlantic Council report. Available from: https://www.atlanticcouncil.org/wp-content/uploads/2022/11/Chinese-Discourse-Power-Ambitions-and-Reality-in-the-Digital-Domain.pdf [Accessed 4 December 2024].

US Department of Defense (2023) 'DoD to establish AI battle labs in EUCOM', 27 September. Available from: https://www.defense.gov/News/Releases/Release/Article/3540283/dod-to-establish-ai-battle-labs-in-eucom-indopacom/#:~:text=Sept.%2027%2C%202023%20%7C%20Two%20BRAVO%20AI%20Battle,from%20Department%20of%20Defense%20%28DOD%29%20operational%20theater%20data [Accessed 4 December 2024].

Xiong, Y. (2018) 'AI军事应用是一把双刃剑' [The military application of AI is a double-edged sword], *People's Liberation Army Daily*, 8 November. Available from: http://military.people.com.cn/n1/2018/1108/c1011-30388886.html [Accessed 4 December 2024].

Xu, X., Kostka, G., and Cao, X. (2022) 'Information control and public support for social credit systems in China', *Journal of Politics*, 84(4): 2230–2245. https://www.journals.uchicago.edu/doi/10.1086/718358

Zeng, J. (2022) *Artificial Intelligence with Chinese Characteristics: National Strategy, Security and Authoritarian Governance*, Dordrecht: Springer.

Zeng, J., Chan, C-H., Schäfer, M.S. (2020) 'Contested Chinese dreams or AI? Public discourse about artificial intelligence on WeChat and People's Daily Online', *Information, Communication & Society*, 25(3): 319–340. https://doi.org/10.1080/1369118X.2020.1776372

Zhan, C. (2018) 'AI技术：未来远超我们的想象' [AI technology: the future will surpass our imagination], *People's Liberation Army Daily*, 30 November. Available from: http://military.people.com.cn/n1/2018/1130/c1011-30434108.html [Accessed 4 December 2024].

Zhang, H., and Wang, L. (2023) 'AI武器化面临的困局' [The dilemma of AI weaponization], *People's Liberation Army Daily*, 9 June. Available from: http://military.people.com.cn/n1/2023/0609/c1011-40009743.html [Accessed 4 December 2024].

Zhang, L. (2023) 'China: generative AI Measures finalized', Law Library of Congress, US Congress, 19 July. Available from: https://www.loc.gov/item/global-legal-monitor/2023-07-18/china-generative-ai-measures-finalized/ [Accessed 4 December 2024].

Zhang, N. (2024) '南非开设国防AI中心' [South Africa establishes a national defence AI centre], *China Military*, 31 May. Available from: http://military.people.com.cn/n1/2024/0531/c1011-40247590.html [Accessed 4 December 2024].

Zhu, J. (2024) 'AI ethics with Chinese characteristics? Concerns and preferred solutions in Chinese academia', *AI & Society*, 39: 1261–1274. https://doi.org/10.1007/s00146-022-01578-w

Zuboff, S. (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, New York: PublicAffairs.

Index

References to figures appear in *italic* type; those in **bold** type refer to tables. References to endnotes show both the page number and the note number (95n1).

S