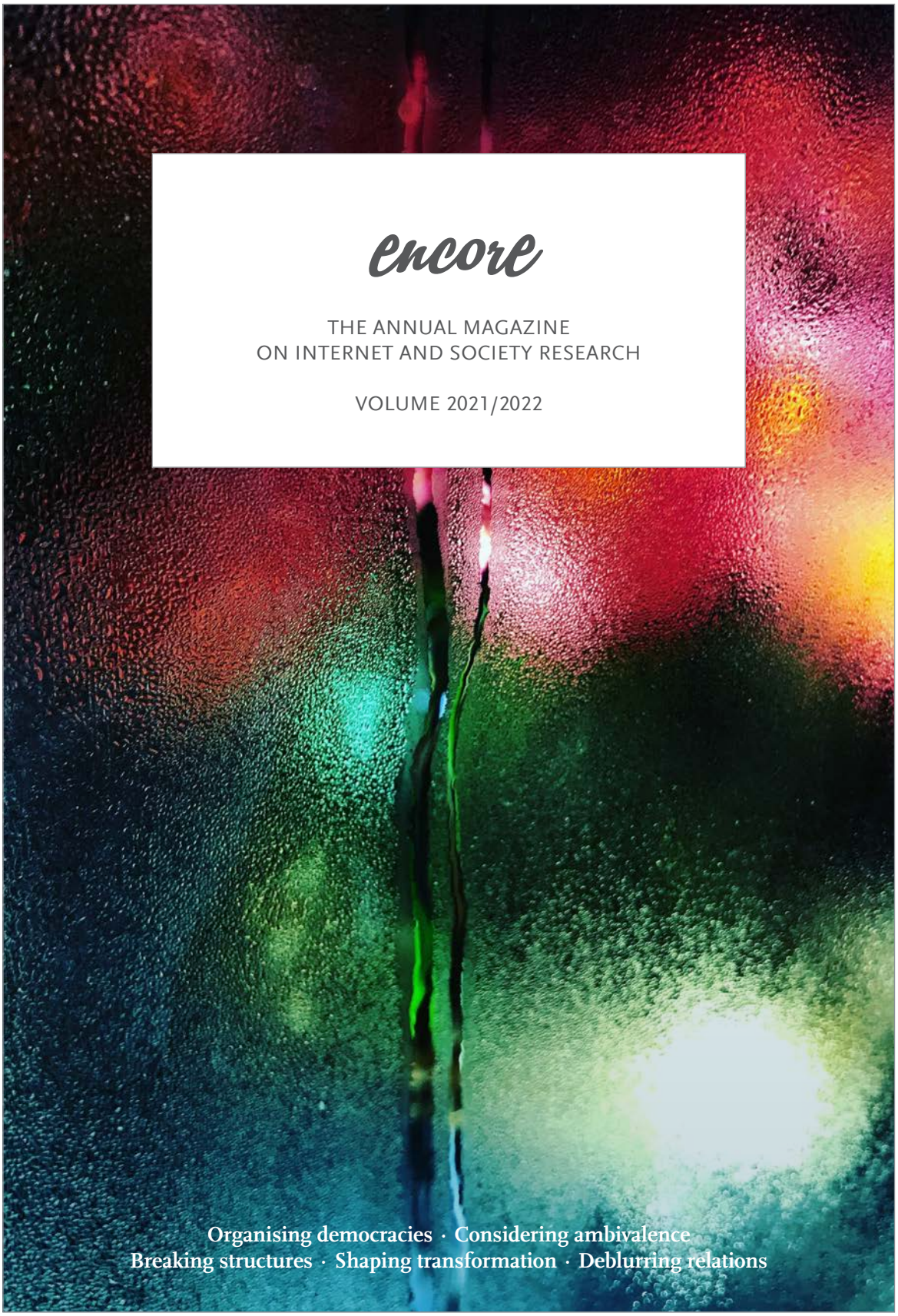


encore

THE ANNUAL MAGAZINE
ON INTERNET AND SOCIETY RESEARCH

VOLUME 2021/2022

Organising democracies · Considering ambivalence
Breaking structures · Shaping transformation · Deblurring relations

encore

THE ANNUAL MAGAZINE
ON INTERNET AND SOCIETY RESEARCH

VOLUME 2021/2022

EDITORIAL

HAPPY BIRTHDAY HIIG AND FULL SPEED AHEAD

Ten years of Alexander von Humboldt Institute for Internet and Society (HIIG), thirty years of the World Wide Web – we have very good reason to celebrate. As scientists, we will leave it up to the readers to decide which of these anniversaries is the most important. We have to acknowledge, though, that one would not be possible in its present form without the other. Many thanks to Tim Berners-Lee and colleagues for their "hypermedia information retrieval initiative aiming to give universal access" to all kinds of information, data and personal contacts to everyone in the world. And of course, our special thanks goes to everyone who has made ten years of internet research at HIIG possible.

In any anniversary year, the experience of looking back is enlightening and exciting. But let's not dwell too long on the past, let's do what we do best at HIIG – analyse, think and design to help shape our digital society.

Social media platforms are celebrating an anniversary too, 15 years of Twitter. Facebook and Google are only slightly older. But the big platforms are no longer carefree teenagers, they have to face increasingly critical questions. That's what happened in the spring of 2021 when Jack Dorsey, Mark Zuckerberg, and Sundar Pichai were questioned before the U.S. Congress about the role of their platforms in spreading hate speech, conspiracy narratives, social segregation, and, in particular, the United States Capitol attack. Prompted by further critique and the outage of Facebook, a discourse has emerged, not only in academia but also in society at large, on how valuable but fragile social platforms have become in our daily interactions. Yes, platforms are essential infrastructure, and this means society as a whole has the task of shaping them. This raises a

multitude of questions about the role of platforms in understanding democracy, about what society actually hopes for from digital platforms and what social function they should assume. The euphoria of the platform teenagers is over; as with every innovation, we don't just need visions but also an awareness of the resulting responsibility. We need an active and sovereign society that not only uses technology but actively shapes it.

Our role as HIIG is that of a researching analyst; we screen developments but also point out alternatives, with the goal of making platforms more democratic, ensuring algorithms are more inclusive and transparent, and integrating principles like intersectional and public interest AI as well as open source.

However, all our analyses, publications, events and stakeholder dialogues over the last 10 years cannot replace political and social negotiation processes but should rather inspire them. This also requires the courage of researchers to create visions of society's future based on scientific expertise. The futures we have envisioned are desirable futures that show how platforms, AI and new work can be designed to function in harmony with the planet's limited resources, in harmony with an inclusive global community that has moved closer together. And let's not forget that through social media, it has never been so easy to knit networks across different continents and cultures. Let's take advantage of the potential that is available to us and continue to shape the future of digitalisation in the next decade.

Congratulations dear HIIG, don't just stay as you are, but keep moving and curious.

Björn Scheuermann,
Director at HIIG

Stephan Bohn,
Project leader at HIIG

CONTENTS

CONSIDERING AMBIVALENCE

BREAKING STRUCTURES

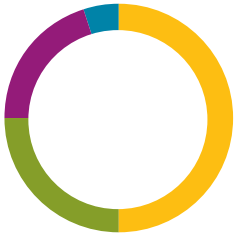
SHAPING TRANSFORMATION

DEBLURRING RELATIONS

ENCORE

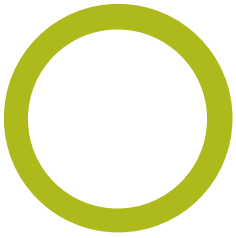
PERCEPTION AND REALITY

What users think a safe password is



- Their username
- Their partner's birthday
- Sequential keys
- Cryptic concatenation of graphemes and punctuation marks

What a safe password really is



- Happy10thBirthdayHIIG!

What crypto art promises to do

- Prep users for life in the metaverse
- Democratise the art market

What crypto art really does

- Enables crypto laundering
- Makes rich artists richer
- Consumes energy
- Allows you to finally own Air Jordans

MARTIN FERTMANN AND MATTHIAS C. KETTEMANN

One council to rule them all: making social media more democratic

Social media companies have become exceedingly powerful. They set rules that influence how online communication takes place. Discussions on how to improve the democratic legitimacy of platform rules have recently gathered pace. Social media councils, as this article shows, can be a powerful tool to bring people and their problems into platform norm-making processes – if they are implemented right.

Parliaments set the rules for democracies. Platforms rule their private online spaces. But as online spaces become ever more important for democratic discourse, we have to ask ourselves: can we make platforms more democratic? We believe social media councils may be the solution to make platform rule making and rule enforcement more accountable, transparent and legitimate.

Few decisions were as highly anticipated in the platform research community as the decision of Facebook's Oversight Board on the question of whether Donald Trump should be allowed to post on the platform again. In the end, the Oversight Board tossed the ball back into Facebook's court and told the platform to take another look at the Trump ban and, on this occasion, clarify their rules and the intended sanctions. In addition, the board members told Facebook to investigate what impact the company's own recommendation algorithms and user design had on the increased polarisation of the American public and the events of 6 January 2021, the storming of the US Capitol.

These are all sensible recommendations that are underpinned by social values. Yes, corporations must apply their rules fairly, especially when they are powerful. Yes, companies must not endanger the rights of others, especially with their products. In Germany, it is the courts – as in the *III. Weg* case or more recently in the two connected Federal Court of Justice (Bundesgerichtshof, BGH) cases of 27 July 2021 – or the legislature (for example, through the Network Enforcement Act, NetzDG) that regularly guide platforms to uphold fundamental social and legal values and that oblige them to include human rights and rule of law considerations in applying community standards. But is there a clear added value if a private advisory body like the Oversight Board does so as well?

NOT AN INVENTION OF FACEBOOK

The Oversight Board is not a brainchild of Zuckerberg created out of thin air but an implementation (admittedly à la Facebook) of long-discussed institutional ideas to use bodies that are more responsive to societal demands to legitimise large digital companies' private orders. These institutions, known as social media councils (in German, we call them *Plattformräte*), have been proposed in recent years by associations such as ARTICLE 19 and Global Partners Digital, as well as by the former UN special rapporteur on

freedom of expression, David Kaye, as alternatives to platform-centric rule making and rule execution. They are an attempt to bring the people back into the process of ruling the rules that regulate online speech.

Properly understood, platform councils are not self-regulatory utopias: they should not replace existing models of private and government regulation of social networks but merely complement them, providing additional impetus to improve private rule (enforcement) systems below the threshold of the justiciable.

WHAT CAN PLATFORM COUNCILS ACHIEVE?

While platform councils can help examine possible violations of terms and conditions or community standards in individual cases, their true added value lies in their capacity to systemically improve companies' governance systems beyond individual cases.

Platform councils can formulate requirements for terms of service, enforcement practices and algorithms, and review and complaint procedures that go beyond individual cases and generate implementation pressure on companies, including through public criticism. In the process, they can encourage institutional change towards more accountable internal normative orders.

HOW SHOULD PLATFORM COUNCILS BE STRUCTURED?

As of now, we do not have enough reliable empirical data on the design of platform councils. At present, it seems that the optimal solution would combine a (quasi-judicial) complaints body and (quasi-legislative) participation in shaping the rules, as this would enable systemic improvements and remedies in specific cases.

In view of the legitimacy deficit of private standard-setting, one opportunity would be to engage in democratic experiments that might entail staffing the councils with randomly selected users or citizens as democratic *mini-publics*. Yet, their lack of expertise might counsel against this. A panel of experts might not be as representative as a mini-public, but their epistemic legitimacy might make up for this. For the most part, human rights experts are very good at setting rules. Unlike in the case of the Oversight Board, an industry-wide body has the effect

of preventing structural dependencies on the companies to be overseen, but individual boards or councils have the advantage of working more quickly. An important bellwether in this regard will be whether we will see other companies joining the Oversight Board.

Social media councils do not have to reinvent the wheel: well-known forms of non-state media supervision by industry-wide self-regulatory bodies such as press and advertising councils can serve as a source of inspiration. Experience with the supervisory bodies for the state media authorities or the broadcasting councils of public broadcasters should also be used to design platform councils, but they should not be applied schematically because of the greatly differing control requirements.

WHAT ARE THE DANGERS OF PLATFORM COUNCILS?

The suspicion voiced by activists that such councils merely shift responsibility is understandable: if inadequately designed, platform councils run the risk of concealing actual power structures without initiating real change. They must therefore not only meet high transparency requirements but must also be equipped with information rights and linked to data access initiatives, so that different actors can understand the extent to which recommended changes actually occur. In its current form, the Facebook Oversight Board has very limited resources in this regard. Whether it can successfully monitor the implementation of its decisions and recommendations in the future depends largely on its willingness to become a powerful actor – to self-constitionalise, in a way. We see some of this happening. For instance, the board told Facebook that it could not just withdraw a case already under consideration, and that it would definitely also consider algorithmic monitoring and content governance as part of the overall rule-enforcement package. However, it noted, it would not do the work of deciding on whether or not to let Trump back online.

WHAT REQUIREMENTS SHOULD BE PLACED ON PLATFORM COUNCILS?

The discussion on platform councils is closely intertwined with demands that platforms should align their often private rules with international human rights

standards. At least insofar as the discussion pursues the idea of an international or regional (European) platform council, these international human rights standards provide a framework not only for informing the decision-making practices of such institutions but also for developing requirements for their design.

If, on the other hand, the concept of a national German platform council finds favour, the decades-long constitutional preoccupation of case law and literature with public broadcasters' supervisory bodies is a normative treasure trove. Such a constitutional reference could, for example, help to balance the independence of the council, which is necessary for effective control, with the cooperative relationship with the company, which is necessary for the effective implementation of its decisions.

WHERE DO WE GO FROM HERE?

Admittedly, platform councils represent only a gradual improvement. However, in the intricate regulatory triangle of states, platform companies and corporations, we can expect little more than gradual improvements. Should platform councils become established, they could resemble their institutional forebears such as the press and broadcasting councils, which have held their own for decades although they have been criticised as compromise solutions.

Despite its shortcomings, the Oversight Board is an important first example of a platform council and provides good material for analysis of both the advantages and disadvantages of such agreements. The board has confidently defined its position in Facebook's regulatory structure and initiated initial changes, not only with its decision in the Trump case but also with its growing body of case law. After an initial flurry of attention, researchers and journalists have started to let the Oversight Board be. We wonder whether now would be the moment to have more rather than less activity, especially as content moderation issues have more recently become a concern for the highest levels of diplomatic governance. To take just one example: How should Facebook react when the Taliban require access to the overthrown government's accounts? And how can Facebook explain to the public that Trump cannot have an account but the Taliban can? (The explanation, by the way, is simple: Trump violated the terms of service numerous times but not all Taliban accounts have consistently done so).

continue reading on page 16 ▸▸

THIS IS AN ARTICLE BY **MARTIN FERTMANN AND MATTHIAS C. KETTEMANN**

This article is based on an op-ed for the *Tagesspiegel Background Digitalisierung & KI* and was first published on 10 May 2021.

Martin Fertmann is a junior researcher at Leibniz Institute for Media Research | Hans-Bredow-Institut (HBI) and holds a PhD fellowship at the Center for Law in Digital Transformation at University of Hamburg. He researches the procedural and institutional implications of international human rights law for platform companies' content moderation practices.

Matthias C. Kettemann is head of the research group Global Constitutionalism and the Internet at Alexander von Humboldt Institute for Internet and Society and scientific co-lead of the Ethics of Digitalisation project. He is Professor of Innovation, Theory and Philosophy of Law at the Institute for Theory and Future of Law, University of Innsbruck, and research program leader at HBI.

At the same time, this specific council should not be exaggerated as an ideal or monopolise the discussion about platform councils terminologically (*oversight boards*) or conceptually. On the contrary, platform councils are a starting point for discussing the future shape of digital governance. ◆

REFERENCES

Kettemann, M. C, & Fertmann, M. (2021). *Die Demokratie plattformfest machen. Social Media Councils als Werkzeug zur gesellschaftlichen Rückbindung der privaten Ordnungen digitaler Plattformen* (Friedrich-Naumann-Stiftung für die Freiheit, 2021/05). https://shop.freiheit.org/#!/Publikation/1055

Fertmann, M., Kettemann, M. C., Morar, D., & Schulz, W. (2021, August 25). Twitter, Trump and the Taliban – the need for better rules for online speech governance. https://leibniz-hbi.de/de/blog/twitter-trump-and-the-taliban

"[I]ncitement capitalism, the sort of insidious combination of surveillance, capitalism and incitement [...] where [people] think they are the victims – and all they are doing is defending themselves."

AN INSIDIOUS COMBINATION OF SURVEILLANCE, CAPITALISM AND INCITEMENT

AN INTERVIEW WITH JAN-WERNER MÜLLER BY THOMAS CHRISTIAN BÄCHLE

Was everything fine with democracy before social media? Global media systems are currently undergoing major structural transformations. How is journalism being affected by these changes? What role do social media platforms and their business models play? Do these developments threaten the proper functioning of representative democracy? Jan-Werner Müller is professor of political theory at Princeton University. After his lecture on *The critical infrastructure of democracy*, our researcher Thomas Christian Bächle invited him for a conversation about our platform-driven media system, the biggest threats to democracies today, free speech on social media, the populists and their business model of incitement capitalism, and the silent majority.

Thomas Christian Bächle: Let's start with a question which might seem to be a bit counter-intuitive at first because it refers to the traditional mass media system rather than social media. The traditional mass media has been said to be an essential power in democracies, the fourth estate. What is your assessment of that notion of a media democracy in the current media environment?

Jan-Werner Müller: Certainly the traditional mass media remains very, very important. The notion that somehow its power has gone, that gatekeepers have been abolished and that the supposedly irrational masses are just waiting for the great demagogue to seduce them is certainly not the case. Especially, but not only in the US, it is clear that at least some of the most outrageous forms of incitement and conspiracy theories could not have caught on the way they have without at least some legitimation from more traditional figures.

They do not have to certify, let alone verify, some of these theories; it is enough that they transmit them and give them some credibility. So, therefore, liberal observers are sometimes wrong to play up the current traditional *demophobia* – the idea that the masses are irrational. They do not see that it is still much more an interplay of traditional institutions and what has indeed become a somewhat unrulier public sphere.

One of your key arguments is that the current controversial debates on social media and platforms are kind of a smokescreen that prevents us from seeing the real problem, which is that the traditional mass media system, in particular local journalism, is in crisis.

It really is an interplay of several developments happening at the same time and maybe sometimes reinforcing each other. Especially in the US, there is a particular kind of press ethos or a particular self-understanding of journalists that runs along the lines of "we have to be neutral and objective and we perform our neutrality and our objectivity by giving both sides equal time, paying equal attention to all our perspectives" and so on. In a well-functioning party democracy, that is a perfectly fine model. But journalists are simultaneously under pressure, from the sense that social media platforms are taking over and that their employment and their business model has become precarious. At the same time, political actors, right-wing populists in particular, constantly tell them that they are biased and if they do not prove that they are really objective and neutral, they are going to feel confirmed that they are "against us".

The problem is not social media in and of itself. It is a particular business model that has siphoned off ads from local journalism. Its destruction makes it more likely that the resulting vacuum will be filled by national news, which, at least in some countries, focuses on cultural war topics and the kind of themes the populists want to talk about. That in turn reinforces the attacks on journalists.

This should also make us think about more holistic approaches. There are obviously plenty of good ideas about how we might fix individual parts, but it is a media system, after all, with many different elements in play. We should focus less on particular people, be they "the people", envisaged as the irrational masses, or Mark Zuckerberg.

What are the biggest threats to democracies today?

The fact that it has become easier to form certain groups and maintain an illusion of immediacy is certainly a threat. So, it's "just me and Trump". Again, it would be fundamentally wrong to say everything was fine with democracy until social media. Of course, we have had plenty of events in the American past somewhat comparable to what we saw on 6 January 2021 with the attack on the US Capitol. To pretend that there were never riots and you could not rile people up, that you could not deploy racism in a particular way would be horrendously naive. But in this particular instance, it became somewhat easier for a figure like Trump to basically practise his political business model. In a sense, it remains available for other actors and clearly can be effective.

This problem in particular is connected to a business model that some of our colleagues rightly call incitement capitalism, the sort of insidious combination of surveillance, capitalism and incitement, which can make it easier to rile people up and manoeuvre them into a particular position where they think they are the victims – and all they are doing is defending themselves. That is very typical rhetoric. You may recall that Trump even went to some of these rallies and said to people: "You are all victims".

That event led to the suspension of Trump's Twitter account. Was that an okay move?

It was okay because it matched the – depending on how you want to use this metaphor – obviously very low American standard of saying: look, free speech is fine until it leads to imminent lawless behaviour. It seemed to be a pretty clear case of somebody continuously inciting, seeing the results of the incitement and still not disavowing it clearly, not really backing off.

And so it is okay for a private company to make that call?

According to American law, it is okay because free speech applies to the government; it does not apply to private companies. In general, I think there are good reasons to criticise this. The sense that free speech is not just a matter of you and the government but basically has an effect on society as a whole is a more plausible understanding of free speech. But legally, it is fine.

For those who think it is not fine, it has something to do with the kind of monopoly power of a very small number of companies. One of the unintended pedagogical effects of what happened in January 2021 is that we learned that all these platforms are basically indispensable for so many smaller players. The fact that you can try to create a rival for Twitter but you cannot find any cloud service because nobody is able to provide the relevant infrastructure is a worrying development. Normally, we would like to think that it should be open and pluralistic, and people should be able to provide us with a new model for having conversations online. But if you basically cannot do it without Google or Amazon, it raises very serious questions about the openness of the public sphere online.

You put forward the idea of infrastructures of democracies. How would that relate to the sort of platform-driven media system that we see today?

There are certain forms of infrastructure that also have the character of a monopoly for the most part. In the past, you basically had one post office. At least with common carriers, it was fine that it was not exactly pluralistic as long as they were neutral and not based on a particular politically relevant business model. People could use at least part of that infrastructure to do what is still essential for a democracy to function well, mainly for individuals to use their communicative freedoms, such as free speech, free assembly and so on. Now the world looks a bit different.

Given that some of today's platforms clearly do not understand themselves ultimately as common carriers, nor are they treated as such by the authorities, we have a problem. One approach would be to ensure a more systemic form of pluralism. Another would be to say that just as much as we regulate telephone companies, beyond just their break-up, we also made sure they could not discriminate in certain ways – that the algorithms cannot effectively perform the function of segregating people. It would be a mistake to demonise either the inventors of social media and this particular business model or the users. It should make us beware of a kind of defeatism that comes down to saying "fascism equals Facebook", as one of our colleagues put it.

Have the forms and practices of populism changed with the new media environment?

Populists are those who not only criticise elites or the establishment but at the same time claim that only they represent what they typically call the "real people" or "the

silent majority". This has two consequences, both of which are anti-pluralist. First, it means that all other political actors are deemed to be fundamentally illegitimate. And second, people who either disagree with this claim or who do not fit the symbolic construction of the real German, the real American, whatever, are essentially excluded. What matters about populism is not anti-elitism. Any of us can criticise elites, which does not mean we are right, but it is not a threat to democracy. What matters is anti-pluralism, and the business model of populists is always one form or another of exclusion.

I do not agree with those who argue that, for instance, Twitter is an inherently or structurally populist medium. That is too simplistic. It is possible to use it for that purpose, namely the formation of particular collectives who have the illusion of a direct link with this supposedly uniquely authentic representative. It is an affordance, if you like, of that particular technology. I would be very sceptical if people were to say we are seeing a rise of populism because we have these new technological possibilities.

As long as we have representative democracy and actors who can make a particular claim about representation, we will have at least the possibility of populism. The things we now complain about in terms of immediacy were already brought up in the context of radio. The fact that incitement capitalism does what it does is not because of technology. It is because certain people chose that as a business model. All that has a populist dimension, undoubtedly. What I am saying is that it is not inevitable on the basis of these particular forms of media. ◆

The recording of the lecture as well as the full compendium around the lecture series Making Sense of the Digital Society can be found online.

www.hiig.de/digitalsociety

D

SAMER HASSAN AND PRIMAVERA DE FILIPPI

Decentralised autonomous organisation

A decentralised autonomous organisation (DAO) is a blockchain-based system that enables people to coordinate and govern themselves. It is mediated by a set of self-executing rules deployed on a public blockchain, and its governance is decentralised (i.e. independent of central control).

The first references to decentralised autonomous organisation (DAO) emerged in the 1990s and described multi-agent systems in an internet-of-things (IoT) environment (Dilger, 1997) or nonviolent decentralised action in the counter-globalisation social movement (Schneider, 2014). The modern meaning of DAOs can be traced back to the earlier concept of a decentralised autonomous corporation (DAC), coined a few years after the appearance of Bitcoin (Nakamoto, 2008). The term was inherently linked to corporate governance; thus, several alternatives to it appeared, leading to the emergence of decentralised applications (dapps) (Johnston, 2013) and later to the generalisation of DAOs as a replacement for DACs (Buterin, 2014).

While some argue that Bitcoin was effectively the first DAO (Buterin, 2014; Hsieh et al., 2018), the term is currently understood as referring not to a blockchain network in and of itself but rather to organisations deployed as smart contracts on top of an existing blockchain network. The first DAO that attracted widespread attention was a 2016 venture capital fund, confusingly called "TheDAO" (DuPont, 2017). Despite the short life of the experiment, TheDAO has inspired a variety of new DAOs (e.g. MolochDAO, MetaCartel), including several platforms aimed at facilitating DAO deployment with a DAO-as-a-service model, such as Aragon, DAOstack, Colony or DAOhaus. There are multiple coexisting definitions of DAOs in use within the industry. The most relevant are the following:

Buterin, in the Ethereum white paper (Buterin, 2013a, p. 23), defines a DAO as a "virtual entity that has a certain set of members or shareholders which [...] have the right to spend the entity's funds and modify its code".

Some of the most popular DAO platforms, such as DAOstack and Aragon, define a DAO similarly as "a network of stakeholders with no central governing body" (https://daostack.io), "which is regulated by a set of automatically enforceable rules on a public blockchain" (https://aragon.org/dao).

In the academic literature, multiple attempts have been made at providing a specific definition of DAOs. Most of these definitions include the following distinctive characteristics:

- DAOs enable people to coordinate and self-govern themselves online.
- A DAO source code is deployed in a (public) blockchain with smart contract capabilities like Ethereum.
- A DAO's smart contract code specifies the rules for interaction among people.
- Since these rules are defined using smart contracts, they are self-executed independently of the will of the parties.
- DAO governance should remain independent of central control.
- Since they rely on a blockchain, DAOs inherit some of its properties, such as transparency, cryptographic security, and decentralisation.

A DAO is not a particular type of business model or a particular type of organisation but a concept that can be used to refer to a wide variety of things. For example, a DAO can be used to create a virtual entity that operates as a crowd-funding platform, a ride-sharing platform, a fully automated company or a fully automated decision-making apparatus.

Scholars have recently started investigating the possibilities for blockchain technology and smart contracts to experiment with open and distributed governance structures (Leonhard, 2017; Rozas et al., 2018), along with the challenges and limitations of doing so (DuPont, 2017; Verstreate, 2018). There is also an emerging body of literature from the field of economic and legal theory concerning DAOs. While most of these works focus on the new opportunities offered by decentralised blockchain-based organisations in the realm of economics and governance (Davidson et al., 2016, 2018), others focus on the legal issues of DAOs, either from a theoretical (De Filippi & Wright, 2018; Reijers et al.,

2018) or practical perspective (Rodrigues, 2018; Werbach, 2018). Various authors have pointed out that DAOs could be used to further economic and political decentralisation in ways that may enable a more democratic and participatory form of governance (Atzori, 2015; Allen et al., 2017). However, in the aftermath of the TheDAO hack, the limitations of blockchain-based governance came to light (DuPont, 2017; Reijers et al., 2018).

The use of the term decentralised autonomous organisation or DAO is now fairly established in the blockchain space, yet there are still many misconceptions and unresolved issues in the discussion around the term. More specifically, there are debates regarding 1) the degree and level of decentralisation, i.e. whether decentralisation is sufficient at the infrastructural level or if it is also necessary at the governance level; 2) the level of autonomy and automation required for its governance, i.e. whether the DAO should operate with or without any human intervention; 3) the conditions required for a community of actors interacting with a smart contract to be regarded as an actual *organisation*. The discussion on whether a DAO should be recognised as a legal person has important implications in the legal field; the common understanding today is that the *autonomous* nature of a DAO is incompatible with the notion of legal personhood, as no identifiable actor is responsible for the actions of the DAO. ◆

The *Glossary of decentralised technosocial systems*, edited by Valeria Ferrari for Internet Policy Review, is an interdisciplinary glossary on peer-to-peer, user-centric and privacy-enhancing decentralised technologies. In order to tackle the existing gap in shared semantics, this glossary converges the efforts of experts from various disciplines to build a shared vocabulary on the social, technical, economic, political aspects of decentralised, distributed or sovereign technologies.

policyreview.info/glossary

REFERENCES

Allen, D. W., Berg, C., Lane, A. M., & Potts, J. (2017). The economics of crypto-democracy. *26th International Joint Conference on Artificial Intelligence, Melbourne*. https://doi.org/10.2139/ssrn.2973050

Atzori, M. (2015). Blockchain technology and decentralized governance: Is the state still necessary? https://doi.org/10.2139/ssrn.2709713

Buterin, V. (2013a). Ethereum whitepaper: A next-generation smart contract and decentralized application platform. https://blockchainlab.com/pdf/Ethereum_white_paper-a_next_generation_smart_contract_and_decentralized_application_platform-vitalik-buterin.pdf

Buterin, V. (2014, May 6). DAOs, DACs, DAs and More: An Incomplete Terminology Guide. Ethereum Foundation Blog. https://blog.ethereum.org/2014/05/06/daos-dacs-das-and-more-an-incomplete-terminology-guide

Davidson, S., De Filippi, P., & Potts, J. (2016a). Disrupting governance: The new institutional economics of distributed ledger technology. https://dx.doi.org/10.2139/ssrn.2811995

Davidson, S., De Filippi, P., & Potts, J. (2016b). Economics of Blockchain. https://doi.org/10.2139/ssrn.2744751

Davidson, S., De Filippi, P., & Potts, J. (2018). Blockchains and the economic institutions of capitalism. *Journal of Institutional Economics, 14*(4), 639–658. https://doi.org/10.1017/S1744137417000200

De Filippi, P., & Wright, A. (2018). *Blockchain and the law: The rule of code*. Harvard University Press.

Dilger, W. (1997). Decentralized autonomous organization of the intelligent home according to the principle of the immune system. *1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation*, 351–356. https://doi.org/10.1109/ICSMC.1997.625775

DuPont, Q. (2017). Experiments in algorithmic governance: A history and ethnography of "The DAO," a failed decentralized autonomous organization. In M. Campbell-Verduyn (Ed.), *Bitcoin and Beyond: Cryptocurrencies, Blockchains, and Global Governance* (pp. 157–177). Routledge. https://doi.org/10.4324/9781315211909-8

Hsieh, Y. Y., Vergne, J. P., Anderson, P., Lakhani, K., & Reitzig, M. (2018). Bitcoin and the rise of decentralized autonomous organizations. *Journal of Organization Design, 7*(1), 1–16. https://doi.org/10.1186/s41469-018-0038-1

Johnston, D. (2013). The General Theory of Decentralized Applications, Dapps. *David Johnston CEO*. https://github.com/DavidJohnstonCEO/DecentralizedApplications

Leonhard, R. (2017). Corporate Governance on Ethereum's Blockchain. https://dx.doi.org/10.2139/ssrn.2977522

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. Bitcoin. https://bitcoin.org/bitcoin.pdf

Reijers, W., Wuisman, I., Mannan, M., De Filippi, P., Wray, C., Rae-Looi, V., Vélez, A. C., & Orgad, L. (2018). Now the code runs itself: On-chain and off-chain governance of blockchain technologies. Topoi. https://doi.org/10.1007/s11245-018-9626-5

Rodrigues, U. R. (2018). Law and the Blockchain. *Iowa Law Review, 104*, 679.

Rozas, D., Tenorio-Fornés, A., Díaz-Molina, S., & Hassan, S. (2018). When Ostrom Meets Blockchain: Exploring the Potentials of Blockchain for Commons Governance. https://eprints.ucm.es/id/eprint/59643/1/SSRN-id3272329.pdf

Schneider, N. (2014). Are You Ready to Trust a Decentralized Autonomous Organization? Shareable. https://www.shareable.net/are-you-ready-to-trust-a-decentralized-autonomous-organization

Verstraete, M. (2018). The Stakes of Smart Contracts. *Loyola University Chicago Law Journal, 50*(3), 743.

Werbach, K. (2018). Trust, but verify: Why the blockchain needs the law. *Berkeley Technology Law Journal, 33*, 487. https://doi.org/10.15779/Z38H41JM9N

COMMENT

Meddling in politics: an emancipatory compass

In the year 2021, we had what was called a "fateful election". Angela Merkel's tenure as chancellor came to an end and speculations, predictions, wishes, and fears for what and who would come after her were omnipresent. Election fever also gripped the institute – we took it upon ourselves to be a lighthouse for all the sailors on the political sea. In this commentary, Moritz Timm outlines the underlying idea behind the digital electoral compass.

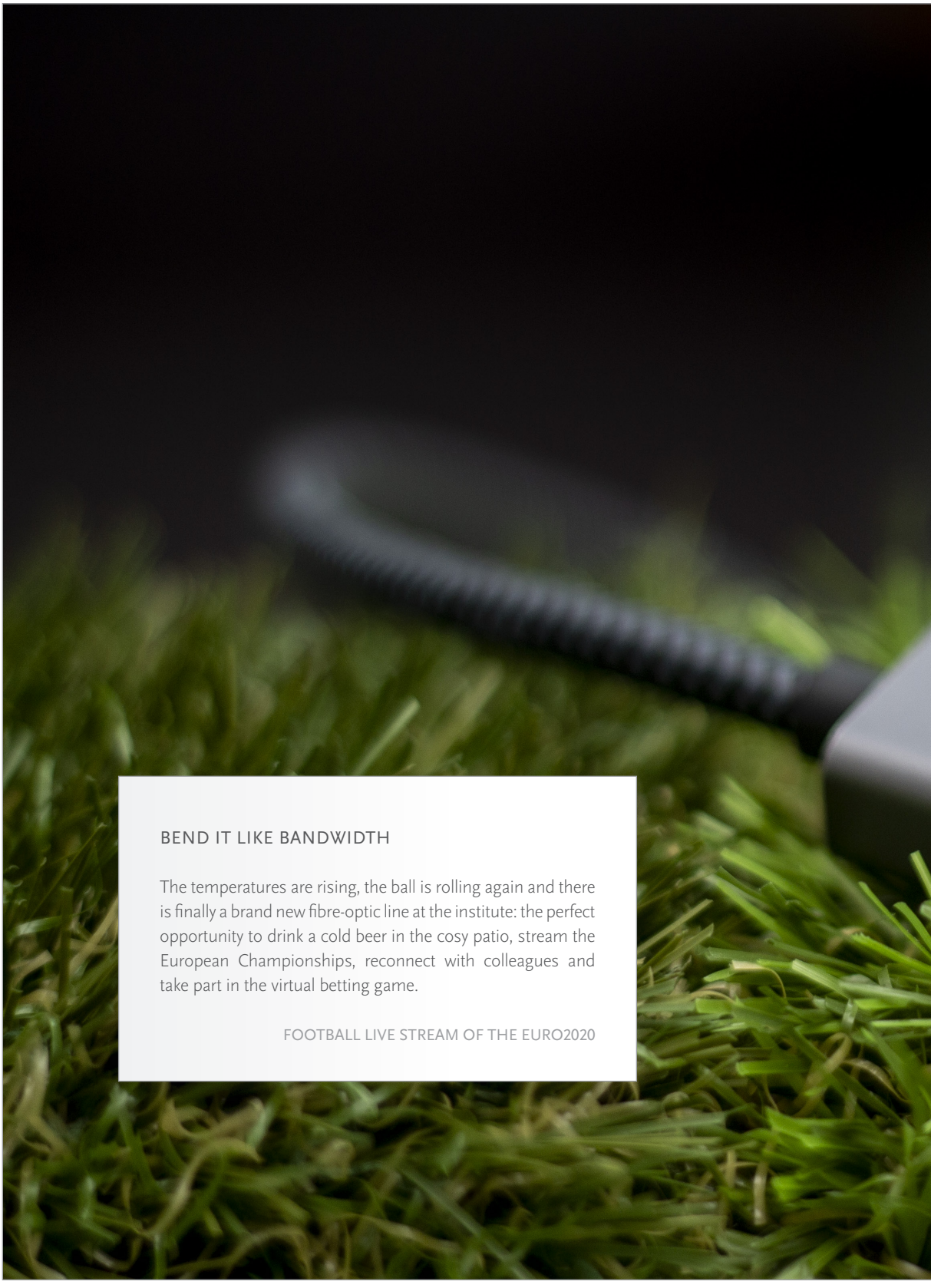
While we do have strong views on several issues, members of our institute usually hold back with their political opinions and statements. As scientists, we prefer hard data, well theorised statements and peer-reviewed findings. We thus aim to inform rather than spinning things politically. Especially because we enjoy finding things out on our own, we found a way to contribute to this year's general election while still remaining impartial. People who read for a living usually know two things: it is fun and can get very, very tiresome in certain instances. Reading election manifestos usually fits into the latter category: it can get tedious. The documents are long, repetitive and full of dull phrases, general statements and very few concrete assessments. They are, after all, textual corporate films, dressed up as action plans.

This might prompt the electorate to disengage. To quote a canonical meme: *Ain't nobody got time for that*. Journalists and press outlets have found a way around that: they do the reading, point out the most important parts and leave you with a somewhat condensed version of the programmes. But they usually do another thing: they spin, they reach conclusions and present them. They are trying to do both the reading and the thinking for their readers. We observed this four years ago and figured that we could act as an intermediary between the electorate and politicians, between spinning for people and leaving them to their own (reading) devices. So we did the reading and the coding – but not the spinning. The electoral compass is what we came up with: it is a low-threshold, open access, easy to use programme scanner. You punch in the keywords, and it shows the statements. It does the finding but not necessarily the reading for you. Whatever conclusions you come up with are your own.

Politically, this is all we are willing to do. We consider the dissemination of knowledge one of our key aims, our *raison d'être* if you will. We sincerely hope that our compass will help you to navigate the convoluted maze of election manifestos and maybe spark people's interest in actually reading the entire thing. Or at least restore faith in the fact that there is actually something hidden in them. You just have to look a little closer. The compass is there to guide your journey. It will, if used correctly, help you sail towards the sun-drenched political atoll of your desires. All hands on deck then, and be weary of the sirens. *Navigare necesse est, vivere non est necesse*, as has been known for two millennia, a statement now possibly more accurate than ever. ◆

Discover the digital electoral compass online.

www.wahlkompass-digitales.de



BEND IT LIKE BANDWIDTH

The temperatures are rising, the ball is rolling again and there is finally a brand new fibre-optic line at the institute: the perfect opportunity to drink a cold beer in the cosy patio, stream the European Championships, reconnect with colleagues and take part in the virtual betting game.

FOOTBALL LIVE STREAM OF THE EURO2020

ANJA DAHLMANN

AI will kill us all: the truth behind the myth

Many sci-fi tropes might be exaggerated, but AI becomes an integral part of military decision-making worldwide. So, AI will help to kill people – but humans must keep control over it.

AI will kill us all! Killer robots will strive for world domination! And invent time travel! While this sci-fi Terminator trope might be a bit over the top, AI is becoming an integral part of military decision-making all over the world. This means that AI will help to kill people because the military applications of AI support novel operational concepts and enable autonomous targeting functions. This accelerates warfare and can lead to better decision making – but also erodes human control, causing legal, ethical and security challenges.

SO HOW DOES AI LEAD TO AUTONOMOUS WEAPON SYSTEMS?

The simple answer here is that, when AI technology is combined with weapon platforms like drones, the resulting systems are autonomous weapon systems (AWS). In the understanding of the International Committee of the Red Cross (ICRC) an AWS is a weapon that "once activated can select and engage targets without further human intervention" (ICRC 2016; similar ICRC 2021, p. 5). This definition is quite broad, but it points us in the right direction.

AI and other technologies enable military users to automate several steps in the targeting process (although AI techniques are not necessary to enable autonomous functions). Such technologies can help to collect big amounts of data and analyse them in almost real time, which minimises the sensor-shooter gap. This can, for example, be applied in the form of pattern of life analysis to create target lists or in image recognition to identify targets. It can also become an important asset in anti-access area denial (A2/AD) operations, which arise when many entities in the battlespace across several domains have to be coordinated to penetrate the adversary's defences. These functions and applications can reach their full potential in a network, where data is collected by all platforms and not just one of them. Some examples include:

The **US Phalanx CIWS** (close-in weapon system), which is installed on ships to detect and destroy incoming missiles or approaching aircraft. This type of defensive, anti-materiel AWS has been used for decades (see Raytheon 2021).

The **Skynet project** has been established by US intelligence agencies to create target lists via behavioural signatures. These are based on mathematical methods using mobile phone metadata. Around 2010, those signatures supposedly identified potential terrorists that became the target of

drone strikes or other sanctions in Pakistan, Yemen and Somalia. Of course, in these cases, human commanders and operators decided on the actual attack. But it does not take much imagination to regard this project as one further step towards an automated targeting cycle to kill humans (see The Intercept 2015a, 2015b).

The Russian **Udar tank** is currently remote controlled, but Russia is working on developing autonomous functions based on advanced sensors and AI techniques and seeking to establish a close connection to other unmanned ground and aerial vehicles (TASS – Russian News Agency, 2021).

The **Future Combat Air System** (FCAS) by France, Germany, and Spain includes a fighter aircraft, and so-called remote carriers – i.e. drones – as well as an IT infrastructure called *combat cloud* and other components employed as a network. The drones might have an option to be armed and might have autonomous targeting functions. The development phase is not yet completed but the plans point in that direction because the drones are supposed to work as *loyal wingmen* to the aircraft. The FCAS is supposed to be deployable by 2040 (Airbus, 2020). The BAE Systems Tempest is a similar project by the United Kingdom and Italy (BAE Systems, 2021).

The **Skyborg Autonomous Core System** is a US project endowed with another sci-fi name that will allow for seamless connectivity of manned and unmanned aerial weapon systems and enable autonomous functions in drones. Those functions include start and landing but also fighting capabilities. It is supposed to be deployable by mid-2023 but will evolve over time (Air Force Research Laboratory, 2021).

This is just a very small selection of worldwide trends and development with a slight bias towards the United States. But even this sample shows that while the concept of automated or autonomous targeting is not all that novel, the technological advances offer a new quality of warfare. The autonomous functions are becoming more flexible in their capacity to support or replace humans in more complex and dynamic situations.

The overall goal of such autonomous targeting functions is to make faster decisions. Minimising communications between humans and machines allows for such faster decisions, because a communication link – as is necessary for remotely-piloted drones – always entails a risk of delay, detection or interruption.

By cutting that out and by being able to analyse huge amounts of data, military decision makers hope to gain a better situational awareness and act faster.

The diverse manifestations of AWS point to problems with simple definitions of the phenomenon – they may not be especially helpful because AWS goes beyond specific, single platforms and also ties in with novel ways of warfare. In this regard, it is worth looking at the human role in the targeting process. The military targeting process consists of formalised steps that vary from military to military in their details and terminology. In the small, dynamic targeting cycle, the steps are find, fix, track, target, engage and assess. These steps directly lead up to the attack (i.e. engage). AI can help to compress these steps due to the fast computing capabilities and minimal communication between human operator and machine. Before this small loop occurs, the wider targeting process includes numerous previous steps – some of which can be performed by AI as well (e.g. Ekelhof & Persi Paoli, 2020).

WHAT ARE THE CONCERNS WITH AI AND AUTONOMOUS FUNCTIONS IN WEAPON SYSTEMS?

So far, we have discussed the beneficial aspects of AI and autonomous functions in the targeting process. But these advantages come at a cost. AI and machine learning processes can be quite opaque for the human operator. Combined with the increasing speed of the targeting process and psychological effects like over-trust and automation, it can become difficult for humans to make meaningful decisions. Machine learning can be executed with data collected before the system is used (offline) or in real-time during the application (online). Especially the latter option can lead to rather uncertain results as it relies on unpredictable environmental data. But even offline learning systems can be unpredictable if they are trained with biased or otherwise erroneous data. And since predictable AWS might be easily fooled by the adversary, militaries might even prefer a certain level of unpredictable behaviour (iPRAW, 2017).

International humanitarian law (IHL) is particularly relevant for the use of autonomous weapon systems. It contains principles like the requirement to distinguish between civilians and combatants (discrimination), the proportionality of means and ends, the military necessity for the use of force and the requirement to take necessary precautions. Whereas some of these issues may already be considered in the run-up to an attack, others must be decided when the actual

situation is underway. The legal accountability for those decisions must remain with the human commander and the weapon system must allow for use when IHL makes this possible. Ultimately, humans remain responsible for killing other people – but the question of when, how, and under which assumptions the decision to pick a specific human target has been made may make a difference. With autonomous functions, planners and programmers will define crucial factors in advance by programming algorithms and picking the training data that will define the indicators for a target. In some cases, and due to the opaqueness of some AI and machine learning techniques, neither them nor the commander or operator might be able to understand the choice of a specific target at a specific time.

This also ties in with ethical concerns, especially with regard to human dignity: algorithms do not understand what it means to kill a human being. "Without this capacity for reflection, however, the human being selected as the target becomes a data point, that is, just an object. The use of autonomous weapon systems would thus violate the dignity of the victim – even technical improvements cannot solve this problem." (Dahlmann & Dickow, 2019, p. 6). From a security perspective, the increasing speed of warfare and the unpredictability of AI decision making could lead to errors and escalations in conflict. Furthermore, new military options will cause new threat perceptions that perpetuate the already ongoing arms dynamics related to military AI (Altmann & Sauer, 2017). This may have detrimental effects on international security and stability.

This discussion of concerns and challenges related to AWS is not meant to romanticise killing and warfare by humans by any means. But the technological advances call for a sober and detailed discussion about tasks and decisions that must remain under human control. Human control encompasses situational understanding by and options for intervention by a human operator – which are enabled by design and exercised during use (iPRAW, 2019). The exact type and level of human control depends on the operational context. For example, deploying a ship with autonomous targeting functions might call for different safeguards than using an armed drone in an urban environment because, for instance, the risk for civilians is much higher in the latter context.

WHAT CAN BE DONE ABOUT THAT?

This is also a question that has not escaped the notice of NGOs and states. Indeed, this is why they initiated a deliberation process at the UN Convention

continue reading on page 40 ▸▸

THIS IS AN ARTICLE BY **ANJA DAHLMANN**

This article was first published on 27 April 2021 on the *Digital Society Blog* of the Alexander von Humboldt Institute for Internet and Society (HIIG). It is based on an episode of HIIG's Why, AI?, an online learning space that unravels myths about automation, algorithms and society curated by Matthias C. Kettemann and Daniela Dicks.

Anja Dahlmann is a researcher and head of the Berlin office of the Institute for Peace Research and Security Policy (IFSH) and a steering group member of the International Panel on the Regulation of Autonomous Weapons (iPRAW).

on Certain Conventional Weapons (CCW). Within this framework, they have been discussing the issue of LAWS since 2014. The deliberations could result in a ban on the development and use of lethal autonomous weapons in the form of a CCW protocol. Other CCW protocols cover incendiary weapons and blinding laser weapons. For both political and conceptual reasons, this outcome does not appear to be particularly likely.

Politically, many states that are capable of building AWS in the near-term future are opposed to a ban, including NATO members to varying degrees. Conceptually, the fact that AWS are not a category of weapon presents a challenge. Instead, this issue is related to functions and the human role in the targeting process. While IHL limits both the means and methods of warfare, this is a rather novel approach (e.g. Rosert & Sauer, 2020).

There might, however, be little room for the regulation of AWS in the form of an obligation to maintain human control: a majority of states at the CCW share the view that humans must keep some sort of oversight, judgment or control. Building on this common belief, states might be able to create a new legally binding treaty or at least a political declaration that sets a normative frame.

In addition to the CCW, there are other fora discussing AWS and the military use of AI – for example, the European Union (see European Parliament 2021, also in the European Defence Fund) and NATO. Furthermore, some states are creating national legislation around AWS. However, this evolving normative framework is progressing rather slowly, which means that the deliberative norm-making process is lagging behind the technological development, leaving it to militaries and industry (Bode, 2019).

But even if the CCW creates a multilateral regulation of AWS, AI will (and has already) become an integral part of the targeting process. There may be military benefits to this, but this development must be accompanied critically and cautiously to maintain a sufficient level of human control.

Special thought should be given not only to conventional weapons but also to the combination of autonomous functions and nuclear weapons. Even if the CCW process fails to deliver tangible results, such a regulation would be worth pursuing, as technical failures, decreases in international stability and increases in the speed of nuclear warfare would have catastrophic effects that may in fact kill us all. ◆

REFERENCES

Air Force Research Laboratory. (2021). Skyborg. https://afresearchlab.com/technology/vanguards/successstories/skyborg

Airbus. (2020). Future combat air system: Owning the sky with the next generation weapons system. https://www.airbus.com/newsroom/stories/Future-Combat-Air-System-Owning-the-sky-with-the-Next-Generation-Weapons-System.html

Altmann, J., & Sauer, F. (2017). Autonomous weapon systems and strategic stability. *Survival, 59*(5), 117–142.

BAE Systems. (2021). The future of combat air: a national endeavour. https://www.baesystems.com/en/the-future-of-combat-air

Bode, I. (2019). Norm-making and the global south: Attempts to regulate lethal autonomous weapons systems. *Global Policy, 10*(3), 359–364. https://doi.org/10.1111/1758-5899.12684

Dahlmann, A., & Dickow, M. (2019). Preventive Regulation of Autonomous Weapon Systems: Need for Action by Germany at Various Levels. https://doi.org/10.18449/2019RP03

Ekelhof, M., & Persi Paoli, G. (March 2020). The Human Element in Decisions about the Use of Force. UNIDIR. https://unidir.org/sites/default/files/2020-03/UNIDIR_Iceberg_SinglePages_web.pdf

European Parliament. (2021). European Parliament resolution of 20 January 2021 on artificial intelligence: questions of interpretation and application of international law in so far as the EU is affected in the areas of civil and military uses and of state authority outside the scope of criminal justice (2020/2013(INI)): P9TA-PROV(2021)0009. https://www.europarl.europa.eu/doceo/document/TA-9-2021-0009_EN.pdf

ICRC. (April 2016). Views of the International Committee of the Red Cross (ICRC) on Autonomous Weapon Systems. https://www.icrc.org/en/document/views-icrc-autonomous-weapon-system

ICRC. (2021). ICRC Position on Autonomous Weapons. https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems

The Intercept. (May 2015a). SKYNET: Applying Advanced Cloud-based Behavior Analytics. https://theintercept.com/document/2015/05/08/skynet-applying-advanced-cloud-based-behavior-analytics

The Intercept. (May 2015b). SKYNET: Courier Detection via Machine Learning. https://theintercept.com/document/2015/05/08/skynet-courier

iPRAW. (November 2017). Focus on Computational Methods in the Context of LAWS. https://www.ipraw.org/wp-content/uploads/2017/11/2017-11-10_iPRAW_Focus-On-Report-2.pdf

iPRAW. (August 2019). Focus on Human Control. Berlin. https://www.ipraw.org/wp-content/uploads/2019/08/2019-08-09_iPRAW_HumanControl.pdf

Raytheon. (2021). Phalanx weapon system. https://www.raytheonmissilesanddefense.com/capabilities/products/phalanx-close-in-weapon-system

Rosert, E., & Sauer, F. (2020). How (not) to stop the killer robots: A comparative analysis of humanitarian disarmament campaign strategies. *Contemporary Security Policy*, 1–26. https://doi.org/10.1080/13523260.2020.1771508

TASS – Russian News Agency (2021, February 11). Russia's latest udar robot to learn to fight on its own and interact with drones. https://tass.com/defense/1255189


BUSTED
MYTH

THOMAS CHRISTIAN BÄCHLE

Deep fakes: the uncanniest iteration of manipulated media content so far

In an age when we so heavily rely on the media for our sense of reality and our idea of the self, a phenomenon such as deep fakes is bound to evoke unease: a perfect illusion, projecting the strange into what is most ordinary. Deep fakes are certainly not the first occurrence of manipulated media content, so what fuels this extraordinary feeling of uncanniness we associate with them?

In early 2021, fake videos featuring a bogus version of actor Tom Cruise circulated on social media alongside comments that either praised their quality or lamented their worrisome perfection. These videos involved whole body movements (including Cruise's characteristic mannerisms), which were performed by another actor for this purpose. At first sight, the short-lived attention given to the counterfeit Tom merely illustrates that deep fakes are becoming increasingly common. So why was this collectively regarded as something noteworthy at all? The answer probably has to do with the performance in the videos, which broke with the expected deep fake aesthetics that usually focuses on the face alone. A whole body copy of a well-known individual, however, represented another giant step towards an unrecognisable and hence troubling illusion.

MANIPULATED MEDIA CONTENT IS A HISTORICALLY CONSISTENT OCCURRENCE

Despite the impression one can easily get from these perfect simulations, practices of manipulating persuasive content in different media forms have a long tradition, of which deep fakes are only one of the most recent iterations. More generally, the term refers to the use of machine learning (deep learning) to create simulated content that pretends to give a truthful depiction of the face, and sometimes also of the voice, of a real person. The increasing relevance of deep fakes, at least within public and political discourses, comes from the combination of the relatively inexpensive access to these technologies, which has coincided with a discomforting rise of misinformation campaigns on video sharing sites and social media platforms. Deep fakes are commonly associated with the communicative intention to deceive and to potentially manipulate. They raise concerns about personal rights or the consequences for mediated realities, including public discourses, journalism and democratic processes. In essence, they are seen by some as nothing less than a "looming challenge for privacy, democracy, and national security" (Chesney & Citron, 2019). Interpretations of these developments often include two very familiar tropes: dramatic accounts of what is new and dangerous about an evolving technology (e.g. Greengard, 2020) are put into perspective by those approaches, which emphasise that there is "nothing new here" and instead demand a shift in focus towards the underlying social structures (Burkell & Gosse, 2019). Between these two perspectives, which are tilted either towards technological determinism

or social constructionism, it is necessary to find middle ground. This middle ground should take into account the recurring motifs commonly associated with the rise of any (media) technology, but at the same time emphasise some defining characteristics of deep fakes that make them a powerful tool of deception.

In other words, it comes down to the tentatively voiced question of what is actually new about the deep fake phenomenon. Taking a historical perspective, it is pretty obvious that the practices of manipulating content in order to influence public opinion far predates digital technologies. Attempts to manipulate images are as old as photography itself. With the most recent technological era, the critical stance on simulated content is even fundamentally ingrained in debates on digital media, with simulation being one of their core characteristics. Questions on the loss of authenticity and auctorial authority – similar to the current anxieties voiced around deep fakes – were also habitually raised with previous media technologies, digital photography in particular (Lister, 2004). When stylising the digital photo, with all its possibilities for easily altering each pixel independently and thus bringing it closer to a painting than a representation of reality, the dubitative, the profound and inescapable doubt of what we see in it, lies at the core of its aesthetics (Lunenfeld, 2000).

At the same time, and irrespective of these oftentimes sinister undertones of manipulation, various types of computer-generated imagery (CGI) have long been applied in the creative industries, providing elaborate visuals in films (Bode, 2017), so-called photo realism in virtual reality environments or life-like avatars in computer games by means of performance capture technologies (Bollmer, 2019). These historical predecessors of doctored content and simulation aesthetics resonate well with the idea of historical continuity and contradict a stance that regards deep fakes as a major disruption. While these analogies certainly have a point, they also have a tendency to disregard today's radically different media environments, in particular their fragmentation (e. g. Poell, Nieborg, & van Dijck, 2019).

SO WHAT'S NEW ABOUT THE DEEP FAKE PHENOMENON?

When addressing the question of what makes deep fakes different from previous media phenomena, one could point to a combination of three factors. The first is the aforementioned fragmented media environment, which is the direct result of the business models social media platforms thrive on. Their consequences

continue reading on page 48 ▸▸

THIS IS AN ARTICLE BY **THOMAS CHRISTIAN BÄCHLE**

This article was first published on 18 January 2022 on the *Digital Society Blog* of the Alexander von Humboldt Institute for Internet and Society (HIIG).

Thomas Christian Bächle is head of the research programme The Evolving Digital Society at HIIG. In 2020 he was JSPS fellow at the Cognitive Science Lab at Waseda University, Tokyo, and 2019/20 guest professor for Media Studies/Digital Media at Humboldt-Universität zu Berlin. His areas of research include human-machine interaction, interfaces and cultural representations of identities, bodies and media technologies.

are felt in traditional journalism, which has been drained financially, as well as in an increasing formation of mini publics that are even reduced to personalised feeds that often lack proper fact checking. This does not just make it easier for misinformation to spread online; the personalised content also allows a form of communication that is often shaped by a high degree of emotionalisation, with the potential to incite groups or individuals.

The second and third elements both cater to a specific aesthetic and only form an effective bond in combination: the suggestive power of audio-visual media and the moving image – still the one media form that supposedly offers the strongest representation of reality – is paired with the affective dimension of communication to which the human face is central. In other words, deep fakes evoke an extraordinary suggestive power by simulating human faces in action. Text-based media are usually met with a more critical distance by media literate readers or users – an awareness that increasingly extends to social media platforms and video sharing sites. This degree of media literacy, however, is challenged by the depiction of faces, especially those that are already known from other contexts. They are central to affective modes of communication and give pre-reflexive cues about emotional and mental states. The human face can even be regarded as the prime site of qualities such as trust and empathy, conveying truth and authenticity despite the cultural differences in how they and the emotions they convey are represented and interpreted.

DEEP FAKES AND OUR SENSE OF REALITY

This somatic dimension clearly hints at a complex relationship between technology, affect and emotion. Unsurprisingly, for individuals, some of the most feared consequences relate precisely to these affective and somatic dimensions of the technology, which can be directly linked to the fake content that is provided. A person's real face and voice, for example, can be integrated into pornographic videos, evoking real feelings of being violated, humiliated, scared or ashamed (Chesney and Citron 2019, p. 1773). In fact, most deep fake content is pornographic (Ajder et al., 2019). This complex relationship between technology, affect and emotion gains even more relevance when we consider the fact that the content provided by and the interactions facilitated on social networks are increasingly perceived as social reality per se, as part of a highly mediated social life. This is why digital images and videos affect both the individual and private idea of the self and the social persona of the public self (McNeill, 2012). Both are part of a space

that is open, contested and hence, in principle, very vulnerable. Identifying with digital representations of the self can even evoke somatic reactions to *virtual* harm, such as rape or violence that is committed against avatars (cf. Danaher, 2018). It is hardly surprising that the debates around deep fake videos clearly express these enormous social and individual anxieties concerning online reputation and the manipulation of individuals' social personas.

Of course, the suggestive power of these videos bears obvious risks for an already easily excitable public discourse, the textbook example being a fake video of inflammatory remarks by a politician on the eve of election day. By fuelling the fires of uncertainty in our mediated realities, they can easily be seen as exacerbating the fake news problem. The question, however, of why deep fake videos create such considerable unease, exceeds this element of misinformation. It is strongly related to this eerie resemblance to reality that leaves us guessing as to whether or not to trust our senses – an uncanniness, in a Freudian sense, of categorical uncertainty about the strange in the familiar. We are fascinated by the illusions deep fakes create for us; they evoke amusement. At the same time, though, they remind us that our mediated realities can never be trusted at face value. Affecting us on a somatic level, deep fakes make us more susceptible to what they show, but in the end only to urge us to doubt what we actually see.

This is exactly why, contrary to all the grim forebodings, and provided that we tread with caution, this could all actually turn out to be a good thing in today's media environment of competing realities (with some, however, being much more trustworthy than others). It is a lucid reminder of the age-old insight that things are not always what they seem, even though, alas, first impressions deceive many. ◆

REFERENCES

Ajder, H., Patrini, G., Cavalli, F., & Cullen, L. (2019). The State of Deepfakes: Landscape, Threats, and Impact. *Deeptrace*. https://regmedia.co.uk/2019/10/08/deepfake_report.pdf

Bode, L. (2017). *Making believe: Screen performance and special effects in popular cinema*. Rutgers University Press.

Bollmer, G. (2019). The kinesthetic index: video games and the body of motion capture. *InVisible Culture Journal 30*.

Burkell, J., & Gosse, C. (2019). Nothing new here: Emphasizing the social and cultural context of deepfakes. *First Monday. Journal on the Internet, 24*(12). https://doi.org/10.5210/fm.v24i12.10287

Chesney, B., & Citron, D. (2019). Deep fakes: a looming challenge for privacy, democracy, and national security. *California Law Review 107*(6), 1753–1820.

Danaher, J. (2018). The law and ethics of virtual sexual assault. In W. Barfield & M. Blitz (Eds.). *The Law of virtual and augmented reality*. Edward Elgar Publishers, pp. 363–388.

Greengard, S. (2020). Will deepfakes do deep damage? *Communications of the ACM, 63*(1), 17–19. https://doi.org/10.1145/3371409

Lister, M. (2004). Photography in the age of electronic imaging. In L. Wells (Ed.). *Photography: A critical introduction*. Routledge, pp. 295–336.

Lunenfeld, P. (2000). Digital photography: The dubitative image. In P. Lunenfeld (Ed.). *Snap to Grid: A User's Guide to Digital Arts, Media, and Cultures*. MIT Press, pp. 55–69.

McNeill, L. (2012). There is no 'I' in network. Social networking sites and posthuman auto/biography. *Biography 35*(1), 65–82.

Poell, T., Nieborg, D., & van Dijck, J. (2019). Platformisation. *Internet Policy Review, 8*(4). https://doi.org/10.14763/2019.4.1425

"Technology is now part of our lives and it is also essential to democratic processes [...] – so it's especially important to look at it with a view to the public interest."

WITH OPEN MINDS AND OPEN SOURCE: THE DEMOCRATIC POTENTIAL OF PUBLIC-INTEREST-ORIENTED AI

AN INTERVIEW WITH THERESA ZÜGER BY PATRICIA LEU

Technology is increasingly indispensable for democratic processes. However, public and private interests are sometimes at odds here. But what about artificial intelligence (AI)? How can AI be developed in a public good-oriented way, and what does this imply exactly? In an interview by Patricia Leu from Prototype Fund, HIIG researcher Theresa Züger talks about why we need more public good-oriented AI and how we can strengthen it.

Patricia Leu: Theresa, you lead the Public Interest AI research group at the Humboldt Institute for Internet and Society. What exactly is your goal?

Theresa Züger: AI systems are increasingly being implemented in processes of social and political relevance. It is crucial to make sure that AI within the context of sociotechnical systems isn't at odds with democratic principles and society's well-being. How this can be maintained in practice and what it actually means for AI to serve society is still not well defined. Therefore, our overall goal is to develop an understanding of the concept of public interest and translate that understanding into the process of developing artificial intelligence. The question we ask ourselves is: how can this understanding change the process and technical implementation of AI development?

We focus on governance processes. For instance, we look at who is (not) involved when AI is developed and what processes and safeguards are in place. In this way, we try to link AI development back to democratic processes. In our team, PhD students are also working on different prototypes of public interest AI. One prototype uses computer vision to recognise accessibility in images. The goal is to make Wheelmap, an open source app for wheelchair accessible places, even more informative.

Then there is a project in which natural language processing is used to translate German texts into easy-to-read German. For many people, reading German texts is so challenging that they are excluded from participation processes, for example. So there is great democratic potential in this project. A third doctoral student is developing design patterns – such as participatory design – that can be used as manageable standards and generalisable solutions in AI development. This will offer AI developers both general guidelines and concrete tools.

How do you want to approach a definition of public interest?

There is a rich tradition of public interest in political theory that has been evident since ancient times. The topic also finds its way into legal philosophical discourse and the constitutions of some states. We explore the concept in legal philosophy, among other fields, and look at how public interest is negotiated there. The philosopher John Dewey and the political scientist Barry Bozeman are inspiring theorists in this context. Dewey sees democracy as a transformative process and views public interest accordingly: in his view, it cannot be universal but must always be negotiated participatively and deliberately in and by society. The idea of participatory design is central according to this logic. These theses were taken up and developed further by Bozeman. As a research group, we are now taking our cue from this and deriving factors from this theoretical foundation to figure out how AI can be developed for the public interest.

And what have you found out so far?

It is important for me to emphasise once again that a public interest orientation may also manifest itself differently in AI from case to case. In any case, however, it is in contrast to private interests and represents an alternative to purely profit-oriented AI development.

Until now, our research has identified five factors that characterise public good-oriented AI. The first is called justification. Here, we look at why the AI is being developed and deployed, what the goals are, and whether the method – i.e. AI – is really the best one in each case. The second aspect is equality: in the best case scenario, AI should strengthen equality in a society (e.g. through improved accessibility), but in no case should AI weaken it. The third factor is deliberation and participatory design. This is about involving people and working together to understand how AI can work well. Depending on the context, this may be possible and useful in very different forms but it is essential to always think in an inclusive way and communicate transparently

about it. Another point is technical security measures, i.e. whether the software is robust, complies with legal standards and is accurate. The last factor is openness for validation: those affected by the AI decisions, but also third parties, must be able to verify that the sociotechnical system around the AI works, i.e. that the people who work with it also use it for the public interest. For that reason, it is important that AI projects operate on an open source basis as much as possible.

Can you give examples of public interest AI technologies?

That's not so easy to answer, because a lot of projects don't even ask themselves the questions we ask ourselves as a research group. So the data is a bit sparse. We can actually imagine public-good-oriented AI in many areas, for example, in the medical sector or in climate research. A good example of public interest AI that we found is Unicef's Project Connect. Here, satellite data is used to identify schools based on their building shape and then see if they have internet access. If they don't, the project's initiators create one in collaboration with various local stakeholders. Any data sets that are created are published. The project is a great example of how AI can be used to increase efficiency: It would take infinitely longer to search all these places for school buildings on the ground – AI does it in a short time and with few resources.

What is your personal approach to and expectation of public-interest-oriented technology and AI?

My academic background is in political theory as well as media studies. It is my scientific interest and also my mission as a scientist to work in the service of society. The topic "AI for good" is currently on everyone's lips, but the theoretical basis is very thin. For many developers, the technology in the end is rarely really in the public interest. I am interested in looking at this demand for ethical and public interest AI with a sound concept and developing a manageable benchmark for it. I strongly believe that public interest AI is possible, but much more work needs to be put into it. Technology is now part of our lives and it is also essential to democratic processes in my opinion – so it's especially important to look at it with a view to the public interest.

How can we ensure that companies that develop technologies focus on the public interest and follow ethical guidelines?

At the moment, we don't think that's possible because the public interest and private interests are at odds with each other. This could be changed through regulation, but

I think it is unrealistic for technology to be developed only with the public interest in mind. It is much more important to create a space for public interest AI, but also for public-interest-oriented technology in general. It needs an ecosystem with sufficient funding so that developers can afford to work in this area.

In my opinion, an initial incentive should come from the state – for example, to make it easier for public-interest-oriented companies and to create funding opportunities. However, civil society must also play a demanding role, be prepared to enter into an open dialogue and negotiate on an equal footing – which brings us back to John Dewey. The third sector must identify issues of public interest and represent them collectively by creating overarching structures instead of individual foundations and organisations doing their own thing. Scientists can also make a valuable contribution to this. And commercial enterprises can also participate by acting as sponsors. Strengthening AI for the public interest is therefore possible from many different directions! ◆

This interview was first published on 19 July 2021 on *prototypefund.de*.

REFERENCES

UNICEF. (n.d.). *Project Connect*. https://projectconnect.unicef.org/map


docs
proto
src/xcode
.gitignore
CODEOWNERS
CODE_OF_CONDUCT.md
CONTRIBUTING.md
Jenkinsfile
LICENSE
NOTICE

SARAH CISTON AND DANIELA DICKS

Making AI more ethical and easier to understand

How can we make artificial intelligence more inclusive? Intersectional approaches to AI draw on established but marginalised perspectives to help reshape the making and using of AI in fundamental ways. A new Intersectional AI Toolkit collects easy tips.

Intersectional approaches to AI (IAI) are key in ensuring more inclusiveness because they draw on marginalised practices to fundamentally reshape the development and use of AI technologies. Such approaches focus on the question "How is AI changing power?" rather than "Is this technology biased, fair or good?" says Pratyusha Ria Kalluri (2020), founder of the Radical AI Network. Our new toolkit provides an introductory guide to IAI and argues that anyone should be able to understand what AI is and what AI ought to be.

AI BIAS REINFORCES DISCRIMINATION

AI systems have made how some of us work, move and socialise much easier. However, their promises to enhance user experience and provide opportunities have not become a reality for everyone. On the contrary: for many people, AI systems have further widened inequality gaps and exacerbated discrimination instead of tackling these problems at their roots. Even so-called intelligent systems merely reproduce the existing analogue world, including its underlying power structures. This means AI applications – like any technology – are never neutral. Allowing only a small but powerful fraction of society to design and implement AI systems means power imbalances remain or are even amplified by computation. Unfair internet infrastructures will continue to be passed off as impartial ones – and with no one else to say otherwise, we may never be able to imagine it any other way.

WHY WE NEED INCLUSIVE AI

Already marginalised communities are often left out of conversations about what kinds of AI systems should and should not exist, and how they should be created and used – despite the fact that these groups are disproportionately affected by the harmful impacts of AI systems. Scholars like Joy Buolamwini (2016) and 2021 MacArthur fellow Safiya Noble (2018) have cited the dangers of algorithmic injustice in an insidious but widespread set of cases, from shadow banning to predictive policing. With the increasing automation of public and private infrastructures, future AI systems should be made by diverse, interdisciplinary and intersectional communities rather than by a select few. In addition to needing community support in order to address the adverse effects they face, system designers can improve AI for everyone by listening to knowledge gained from many perspectives. Diverse groups

– for example Black feminists and queer and disability theorists – have long been considering aspects of the same questions exacerbated by problematic AI. We can and must rely on a broader variety of perspectives if we are to shift the course of AI toward a future of more inclusive systems.

Building on its research on public interest AI, the AI & Society Lab at Alexander von Humboldt Institute (HIIG) focuses on questions in this area. How can AI and other technologies be made more approachable for everyone to ensure people better understand AI systems and how they affect them? What do particularly marginalised communities wish to change about AI, and how can we support them in doing so?

HOW INTERSECTIONAL AI CAN HELP

The Intersectional AI Toolkit helps answer these questions by connecting communities to create introductory guides to AI from multiple, approachable perspectives. Developed by Sarah Ciston during a virtual fellowship at the AI & Society Lab, the Intersectional AI Toolkit argues that anyone can and should be able to understand what AI is and what AI ought to be.

Intersectionality describes how power operates structurally, and how multiple forms of discrimination have compounding, interdependent effects. American lawyer Kimberlé Crenshaw introduced the term in 1989, using the image of an intersection where paths of power cross to illustrate the interwoven nature of social inequalities (1989). As imagined by this toolkit, intersectional AI will bring decades of work on intersectional ideas, ethics and tactics to bear on the issues of inequality associated with AI. By drawing on established ideas and practices and understanding how to combine them, intersectionality can help reshape AI in fundamental ways. Through its layered, structural approach, Intersectional AI connects the dots between concepts – as seen from different disciplines and operating across systems – so that individuals and researchers may be able to help address the gaps that others could not see.

A TOOLKIT THAT HELPS PEOPLE TO THINK ABOUT INTERSECTIONALITY AND CODE INCLUSIVE AI

The Intersectional AI Toolkit is a collection of small magazines (or zines) that offer practical, accessible guides to both AI and intersectionality. They are written

for engineers, artists, activists, academics, makers and anyone who wants to understand the automated systems that impact them. By sharing key concepts, tactics and resources, they serve as jumping-off points to inspire readers' own further research and conversation across disciplines and communities, asking questions like "Is decolonising AI possible?" or "What does it mean to learn to code?"

The toolkit is available as a digital resource that continues to grow as more community contributions are added, as well as printable zines that can be folded, shared and discussed offline. With issues like a two-sided glossary, "IAI A-to-Z", strategy flashcards, "Tactics for Intersectional AI", and a guide to concepts for sceptics, "Help Me Understand Intersectionality", the zine collection focuses on using plain language and fostering tangible impacts.

This toolkit is not the first or only resource on intersectionality or AI. Instead, it gathers together some of the amazing people, ideas and forces working to re-examine the foundational assumptions built into these technologies, such as Ruja Benjamin's *Race after Technology* (2019) and Catherine D'Ignazio and Lauren Klein's *Data Feminism* (2020). It also looks at what people are (not) involved when AI is developed or what processes and safeguards do or should exist. In this way, it helps us understand power and aims to link AI development back to democratic processes.

WHY IS THE FUTURE OF AI INTERSECTIONAL?

Current approaches to AI fail to address two major problems. First, those who create AI systems – from code to policy to infrastructure – fail to listen to the needs or wisdom of the marginalised communities most injured by those systems. Second, the current language and tools for AI represent intimidating barriers that prevent outsiders from understanding, building or changing these systems. If we want improved, inclusive AI systems, we must consider the needs and knowledge of a broader range of people. Otherwise, we face a future of perpetuating the same problems under the guise of fairness and automation.

The Intersectional AI Toolkit seeks to intervene by facilitating much-needed exchange between different groups around these issues. The AI & Society Lab hosted the launch of the toolkit as an Edit-a-thon workshop in order to gain multiple valuable perspectives through diverse public participation. At the

event, a demo of the toolkit was followed by a lively, interdisciplinary and truly intersectional discussion among artists, activists, coders, and people with no technical knowledge, who were offered a safe space to explore, edit and expand on the toolkit according to their needs. Over the next months, more digital and in-person zine-making workshops are planned to keep building the toolkit while advocating for intersectional approaches to AI in various sectors like AI governance.

All AI systems are socio-technical; they interconnect humans and machines. Intersectionality reminds us how power imbalances affect those connections. By addressing the gap between those who want to understand and shape AI, and those who already make and regulate it, intersectional AI can help us find the shared language we need to reimagine AI together. ◆

REFERENCES

Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity.

Buolamwini, J. (2016, May 16). InCoding — In The Beginning Was The Coded Gaze. *MIT Media Lab*. https://medium.com/mit-media-lab/incoding-in-the-beginning-4e2a5c51a45d

Ciston, S. (n.d.). *Intersectional AI Toolkit*. GitHub. https://intersectionalai.com

Crenshaw, K. (1989). Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics. *University of Chicago Legal Forum, 1989* (1), 139–167. http://chicagounbound.uchicago.edu/uclf/vol1989/iss1/8

D'Ignazio, C., & Klein, L. F. (2020). *Data Feminism*. MIT Press.

Kalluri, P. (2020). Don't ask if artificial intelligence is good or fair, ask how it shifts power. *Nature, 583*(7815), 169–169. https://doi.org/10.1038/d41586-020-02003-2

Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press.

THIS IS AN ARTICLE BY **SARAH CISTON AND DANIELA DICKS**

This article was first published on 26 October 2021 on the *Digital Society Blog* of the Alexander von Humboldt Institute for Internet and Society (HIIG). The Intersectional AI Toolkit was presented and further developed at the Edit-a-thon on 1 September 2021, hosted by HIIG's AI & Society Lab.

Sarah Ciston is a Mellon fellow and PhD candidate in media arts and practice at University of Southern California, where she researches intersectional AI and leads Creative Code Collective – a student community for co-learning programming using approachable, interdisciplinary strategies.

Daniela Dicks was co-lead and spokesperson of HIIG's AI & Society Lab, an inter- and transdisciplinary research laboratory at the intersection of research, industry and civil society.

BERLIN WET HOT SUMMER

Even though the sun barely shined, our hearts warmed at being able to finally meet up with the entire crew. At this year's retreats, we were able to re-engage in real-life interaction and collaborative work. Future walks, idea labs, lightning talks – there were ingenious ideas and provocative research that not only helped us to fill the agenda for our summer retreat but also allowed us to map out the institute's trajectory for the coming years.

SPRING AND SUMMER RETREATS

VICTORIA GUIJARRO SANTOS

It's a match! Or racism?

No technology is neutral. Dating apps like Tinder and Grindr can perpetuate stereotypical assumptions about sexual preferences and reinforce a racist flirting culture. Can the law intervene?

Many dating apps work on roughly the same principle: users are shown profiles of other users. If they find each other attractive, they can contact each other – or as Tinder says: "It's a match". The dating app OkCupid studied the behaviour of its users from 2009 to 2014 and found that Asian men and Black women had the worst chances of finding a date, while white people had the best (Rudder, 2014). Other studies have come to similar conclusions (Callander et al., 2016; Curington et al., 2021). So race seems to be a crucial factor in online dating.

SEXUAL PREFERENCE OR RACISM?

Ideals of beauty and sexual preferences based on them have developed in a specific historical and cultural context that is not free of discrimination (Hutson et al., 2018). As recently as 50 years ago, laws were in force in the USA that prohibited relationships between whites and Blacks in order to ensure white supremacy (US anti-miscegenation law; see also German Nuremberg Laws). Overall, racist legislation has led to the devaluation of non-white people. The figures collected about online dating may indicate that racist power relations continue to influence sexual preferences today (Hutson et al., 2018). Certainly, any person can act contrary to these structures, and closer reflection is needed to identify the conditions under which individual choice of sexual partners based on (un) conscious racist stereotypes is also racist (Bedi, 2015). However, I am not concerned with this individual choice in this text. Instead, the focus is on how dating apps in particular can perpetuate and reinforce racist structures in their function as digital infrastructure.

VALUES IN TECHNOLOGY

It is important to understand that (digital) technology is not a neutral artefact (cf. Winner, 1980). The developers of dating apps have to decide which values to inscribe. How should user interaction be structured? Can statements be deleted? Under what conditions should this happen and by whom? For example, Once, a popular dating app, automatically replaces penis pictures with a cat photo. Only when the recipient agrees does the kitten disappear. This is Once's response to the fact that 51% of its users had been sent unwanted nude pictures on at least one occasion (Once, 2020). Furthermore, questions arise such as: what filter options does a dating app offer? On Grindr, for example, it was possible to filter by "ethnic origin" until recently (Grindr, 2020); on Lovescout24, this is still possible (LoveScout24, 2018).

And then there are the matching algorithms. What criteria should they use to search for matches? What data should be used to develop the algorithmic system? And what role should user feedback play? As a rule, we do not know which matching algorithms are used in each case. What they all have in common, however, is that they want to achieve matches. It is therefore plausible that the respective algorithms show users the profiles of other users according to the probability of success. For example, the "relative attractiveness" of a person is supposed to be decisive in the Tinder algorithm (Online for Love, 2021).

If predominantly white people are found attractive, as shown by the evaluations of OkCupid, among others, then whiteness is a high indicator for a match, i.e. a higher probability of success. So even if skin colour is not an explicit criterion for the algorithmic system (cf. Tinder, 2019), the algorithm can compensate for this missing information via proxy characteristics such as a person's success rate and, as a result, match predominantly white people. In this way, preferences that were once deemed merely personal can be aggregated and become digital infrastructure. Once coded, social biases thus become the program.

POTENTIAL AND SELF-REGULATION

Dating apps may also enable connections between people from different social groups who would otherwise not come in contact with each other (cf. Emerging Technology from the arXiv, 2017; Rosenfeld & Thomas, 2012). However, this potential will be diminished if discrimination is not curbed. Certainly, the companies behind these dating apps have an economic interest in being as inclusive as possible in order to generate more data about their users. This is probably also why Grindr has moved to create an environment "where diversity [...] thrive[s]" (Grindr, 2020, 2021). Tinder also emphasises, "We don't care (or store) whether you're black, white, magenta or blue." (Tinder, 2019). These are positive developments.

But does Tinder also care whether a user is disabled? And how does the company ensure that proxy characteristics do not influence the algorithm? Corporate self-regulation is intransparent and not exhaustive. Most importantly, discrimination is not a new phenomenon. The legal system grants affected persons subjective rights to be protected from discrimination. Their enforcement should not be made dependent on the economic benefit of a company.

ALGORITHMS AND ANTI-DISCRIMINATION LAW

Digital technologies, however, expose the inadequacies of state-set, subjective legal protection and make deficits in law enforcement particularly apparent. This is crystallised in dating apps in many ways: for example, in Germany, the unsolicited sending of penis pictures is a punishable offense (§ 184 of the German Criminal Code); however, it is hardly prosecuted. On the other hand, Grindr and OkCupid sell discrimination-sensitive data to third parties, contrary to the General Data Protection Regulation (Forbruker Rådet, 2020); this was first prosecuted by the Norwegian Data Protection Authority in 2021 (Datatilsynet, 2021). At this point, however, I would like to focus on gaps in anti-discrimination law with regard to matching algorithms.

In Germany, Section 19 of the General Act on Equal Treatment (*Allgemeines Gleichbehandlungsgesetz,* AGG) makes unequal treatment of persons on the basis of the legally designated categories of race, ethnic origin, gender, religion, disability, age or sexual identity in so-called mass transactions subject to a justification requirement. With regard to dating apps, the fact that the matching algorithm processes users' individual data and then shows them aligned profiles of other app users could counter the assumption of a mass business and thus against the application of Section 19. But for the provider, the individual users are not really important. They are one of many whose personal data is processed en masse using the same algorithm; the exact content of the service may vary (profiles displayed) but not the service offered en masse as such (display of profiles) (cf. Hacker, 2018).

If profiles of BIPoC (Black, Indigenous and People of Colour) were systematically displayed less frequently in a dating app than profiles of white users, this could constitute direct or indirect discrimination in the sense of the General Act on Equal Treatment. This would require more detailed investigations in individual cases. Assuming this is the case, however, (at least indirect) discrimination can be justified. This raises questions that are not easy to answer: on the one hand, it is the business model of a dating app to achieve matches, so it is useful to use statistics on sexual preferences. On the other hand, statistics can be biased, reproducing and reinforcing discriminatory stereotypes.

But to what extent should statistical findings then be able to justify discrimination? These questions must first be addressed in legal practice. Knowledge hierarchies

make access to the law more difficult. For one thing, a user has to suspect that she is being discriminated against in the first place. However, since she uses the dating app independently of others, she has no direct comparison. For subsequent legal proceedings, the user would have to prove – based on knowledge she has somehow acquired in support of the presumption – that a legally protected group was discriminated against. Since knowledge hierarchies are not alien to anti-discrimination law, it accommodates plaintiffs by reducing the standard of proof and distributing the burden of proof (§ 22 AGG). For example, in order to prove indirect discrimination, the user would *only* have to prove a discriminatory effect and not the causal discriminatory act, i.e. she does not have to know exactly how the algorithmic system works (Tischbirek, 2020).

Merely demonstrating that BIPoC are treated unequally as a result is enough. The provider must justify why this is so. However, it is simply too much for a single person to compile statistics on the discriminatory user experiences of a particular social group and thus the discriminatory effect. In addition, the user would need a basic understanding of algorithmic systems in order to be able to refute the provider's technical justifications in a court case that might come about despite this. Although the user could commission an expert opinion in order to be on a par with the expert knowledge of the companies in the context of legal proceedings, these are hardly cheap. Affected persons should therefore receive stronger support in legal proceedings and not have to fight discriminatory social structures and tech companies alone (cf. Mangold, 2021; Tischbirek, 2020). To this end, the long-demanded right of associations to sue in anti-discrimination proceedings should finally be introduced in Germany. This is even desirable under EU law (cf. Art. 7 Council Directive 2000/43/EC).

STRENGTHEN LEGAL INSTRUMENTS AND DISCOURSE

Dating apps exemplify all the digital technologies that we use every day and that shape our perceptions. The fact that digital technologies can scale discriminatory structures makes it even more urgent to deal with, name and fight relations of domination in this context as well. To this end, the law must be sharpened and law enforcement improved. Depending on the area of application and the consequences of digital technologies, companies could be subject to graduated obligations. At the very least, however, a broader social discourse is needed on how digital technologies should be designed and how goals critical of domination should be implemented in them. ◆

THIS IS AN ARTICLE BY **VICTORIA GUIJARRO SANTOS**

This article – first published in *Forum Recht 01/2020* – was featured as guest contribution on the *Digital Society Blog* of the Alexander von Humboldt Institute for Internet and Society (HIIG) and is printed here in an abridged and revised form.

Victoria Guijarro Santos is a researcher at the Chair for International Public Law and International Human Rights Law at University of Münster. She is particularly interested in data and law, anti-discrimination law, EU and constitutional law, and feminist legal theory.

REFERENCES

Bedi, S. (2015). Sexual Racism: Intimacy as a Matter of Justice. *The Journal of Politics, 77*(4), 998–1011.

Callander, D., Holt, M., & Newman, C. E. (2016). 'Not Everyone's Gonna Like Me': Accounting for Race and Racism in Sex and Dating Web Services for Gay and Bisexual Men. *Ethnicities, 16*(1), 3–21.

Curington, C. V., Lundquist, J. H., & Lin, K.-H. (2021). *The Dating Divide: Race and Desire in the Era of Online Romance*. University of California Press.

Datatilsynet. (2021). Intention to issue €10 million fine to Grindr LLC. https://www.datatilsynet.no/en/news/2021/intention-to-issue--10-million-fine-to-grindr-llc2

Emerging Technology from the arXiv (2017, October 10). First Evidence That Online Dating Is Changing the Nature of Society. MIT Technology Review. https://www.technologyreview.com/2017/10/10/148701/first-evidence-that-online-dating-is-changing-the-nature-of-society

Forbruker Rådet. (2020, January 14). Out Of Control: How Consumers are Exploited by the Online Advertising Industry. https://fil.forbrukerradet.no/wp-content/uploads/2020/01/2020-01-14-out-of-control-final-version.pdf

Grindr. (2020, June 1). We will not be silent. Black Lives Matter. https://t.co/93ninYt2u9

Grindr. (2021). Community Guidelines. https://www.grindr.com/community-guidelines/?lang=en-US#hate-voilence-offensive-content

Hacker, P. (2018). Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies against Algorithmic Discrimination under EU Law. *Common Market Law Review*, 55, 1143–1186.

Hutson, J., Taft G., J., Barocas, S., & Levy, K. (2018). Debiasing Desire: Addressing Bias & Discrimination on Intimate Platforms. *Proceedings of the ACM on Human-Computer Interaction, 2*(CSCW), 1–18.

LoveScout24. (2018, May 25). Datenschutzbestimmungen. https://www.lovescout24.de/faq

Mangold, A. K. (2021). *Demokratische Inklusion durch Recht: Antidiskriminierungsrecht als Ermöglichungsbedingung der demokratischen Begegnung von Freien und Gleichen*. Mohr Siebeck.

Once. (2020, April 1). April April. https://getonce.com/de/blog/april-april

Online for Love. (2021, April 30). Tinder Elo: the Definitive Guide to the Tinder Algorithm. https://t1p.de/ra7x

Rosenfeld, M. J., & Thomas, R. J. (2012). Searching for a Mate: The Rise of the Internet as a Social Intermediary. *American Sociological Review, 77*(4), 523–547.

Rudder, C. (2014, September 10). Race and Attraction, 2009–2014. https://www.gwern.net/docs/psychology/okcupid/raceandattraction20092014.html

Tinder (2019, March 15). Powering Tinder: Die Methode hinter unserem Matching. https://t1p.de/nayi

Tischbirek, A. (2020). Artificial Intelligence and Discrimination: Discriminating Against Discriminatory Systems. In T. Wischmeyer & T. Rademacher (Eds.), *Regulating Artificial Intelligence* (pp. 103–121). Springer.

Winner, L. (1980). Do Artifacts Have Politics? *Daedalus, 109*(1), 121–136.

KELSEY MEDEIROS

Sexual harassment in academia

This article sheds light on sexual harassment, what role it plays in relation to power structures in academia, and possible ways to address it.

Today is the day Julia has been dreaming about since the first day of her PhD program – it's interview day at the American Economics Association (AAEA) Annual Conference. As her phone plays her familiar morning alarm, Julia jumps out of bed, bright-eyed and hopeful for the 10 interviews she has scheduled for the day. She irons her power skirt and navy blouse carefully, ensuring she leaves no creases uncreased. She reviews her notes on each school, each committee member, and each job requirement, preparing her responses just enough to sound polished but not too much as to sound over rehearsed. She slips on her special interview heels that give her that extra boost of confidence and height she likes and walks out of her hotel room – shutting the door behind her and with it, imagining the start of her future as an assistant professor of economics.

Julia walks straight to her first interview which she finds a few doors down from her own hotel room. After a brief knock, a man opens the door to a hotel room identical to hers. He greets her and she feels the butterflies of meeting a scholar whose research she cites regularly and who is considered a giant in the field of economics. She has admired his work since early in her PhD training and has dreamed about the chance to discuss it with him.

As she enters, she scans the room and notices one chair which has been turned from the hotel desk to face the edge of the bed. The committee member takes his seat and gestures toward the bed. Julia had read about women who had been asked to sit on the hotel bed, and some even asked to lie down, during their job interviews. She also read that some had been verbally and physically propositioned during their interviews! The stories were legendary whispers in her women in economics networking groups. But that was the "old days" she had told herself. She hadn't expected to find herself in this compromising position in 2018. #Metoo had been making headlines for some time now and university faculty would have surely been paying attention.

She took her seat on the bed. She instantly felt uncomfortable as she sat there, the man staring at her from his chair. Her thoughts became scrambled as she thought about the stories she had read about and whether or not the same might happen to her – What will I do? How can I get out? I can't make a bad impression on this guy – this is my career. Breathe Julia.

She misses the interviewer's question. "I'm sorry, could you repeat that?" she asks. Looking slightly annoyed, the committee member repeats his question, but Julia still struggles to focus. She responds to the question, but fails to make her point clearly, jumbling her words and forgetting key details. The mental notes she knew so well just minutes ago in her own hotel room, now seem buried under anxieties and fear for her own wellbeing.

Some may read Julia's story and think she should have just rolled with the punches – a bed is just a big, squishy chair with pillows. If anything, she should feel thankful that she was given such a throne for her interview. For others, the power dynamics in Julia's story are palpable. Many will see that what some may construe as a harmless situation, could quickly turn harmful. Indeed, stories and research on power dynamics in academia reveal that situations like Julia's can and do turn precarious, at times resulting in sexual harassment and even assault.

Sexual harassment on its own is problematic and worth addressing. However, the issue compounds when one considers who the typical target of sexual harassment is. Research tells us that individuals who are lower in a status hierarchy and those that are unrepresented are more likely to be the targets of harassment behavior (e.g., Harned et al., 2002). In the academy, this means that women in, or in contention for, junior faculty positions are disproportionately targeted when it comes to sexual harassment. From an intersectional lens, BIPOC and LGBTQIA+ individuals in the academy may be particularly at risk of sexual harassment.

The research on sexual harassment and career trajectory is clear – when someone is harassed, they are more likely to leave an organization and with some leaving the field altogether (McLaughlin et al, 2017; Medeiros & Griffith, 2019). By failing to address sexual harassment among faculty in our institutions, we are systematically pushing these groups out of our institutions. We are pushing bright minds, new ideas, and critical perspectives out of our classrooms and out of our labs. In a time when universities are increasingly committing to creating inclusive spaces, they are undermining their efforts by not addressing the systemic issues within academia that perpetuate harassment.

IS SEXUAL HARASSMENT IN THE ACADEMY A PROBLEM?

Yes. Research on sexual harassment in the academy suggests that it remains a prevalent problem. In a 2003 study examining incidences of sexual harassment in the workplace across private, public, academic, and military industries, Ilies et al (2003) found academia to have the second highest rates of harassment, second only to the military. More recently, a report by the The National Academies of Sciences, Engineering, and Medicine (NASEM) summarized the persistent problem of sexual harassment in academia with regard to faculty-student harassment, as well as faculty-faculty harassment. To find more evidence of this issue, one can also turn to Twitter – as Times Higher Education highlighted in their 2019 blog.

There are several systemic reasons why sexual harassment may be particularly prevalent in academia. As recognized by the NASEM's 2018 report on sexual harassment in the academy, these include academia's gender imbalance and its hierarchical power structure. Research suggests that sexual harassment is more likely to occur in male-dominated organizations (e.g., Hegewisch & O'Farrell, 2015; Medeiros & Griffith, 2019). Although academia has made advances with regard to increasing the number of women in faculty positions, it largely remains a male-dominated industry, especially in more senior faculty positions (Bacchi, 1993; Diamond et al., 2016; O'Connor, 2020; Zhuge et al., 2011). Additionally, the hierarchical nature of academia, as well as its over emphasis on "star researchers" creates an environment in which sexual harassment by those in power may perpetuate.

Along these lines, safe reporting mechanisms must be put in place to encourage junior faculty members to bring forward issues that challenge the extant hierarchy. This requires clear policies and the communication of these policies. Research from business, however, also teaches us that comprehensive protections must be in place for those who do report. For example, in an experimental study, Hart (2019) found that women who self reported sexual harassment were less likely to be recommended for promotion compared to women with identical qualifications. The failure to set up systems that protect those that report sexual harassment and challenge academia's hierarchy, enables a self-perpateuating cycle of power to flourish. Is it really such a surprise then that researchers such as Kirkner, Lorenz, and Mazar (2020) found that sexual harassment largely goes unreported?

WHAT CONSTITUTES SEXUAL HARASSMENT?

Definitions of sexual harassment are similar across borders generally including both verbal and physical manifestations. See below for examples of how it is defined around the world:

The **US Equal Employment Opportunity Commission (EEOC)** defines sexual harassment as follows: "unwelcome sexual advances, requests for sexual favors, and other verbal or physical conduct of a sexual nature."

The **UK government** defines sexual harassment as: "any unwanted conduct of sexual nature that makes you feel intimidated, degraded, humiliated, or offended."

In **Japan**, harassment is defined more broadly as "power harassment" or "pawa hara," which includes: "physical abuse, emotional abuse deliberately isolating an employee, overworking an employee, consistently assigning work below an employee's skill level, and infringing on an employee's privacy.

The **South African government** lists the following as their definition of sexual harassment: "(1) Sexual harassment is unwanted conduct of a sexual nature. The unwanted nature of sexual harassment distinguishes it from behaviour that is welcome and mutual. (2) Sexual attention becomes sexual harassment if: (a) The behaviour is persisted in, although a single incident of harassment can constitute sexual harassment; and/or (b) The recipient has made it clear that the behaviour is considered offensive; and/or (c) The perpetrator should have known that the behaviour is regarded as unacceptable"

Cross culturally, it appears we can agree – sexual harassment invovles unwanted verbal and phsyical abuse.

WHY POWER MATTERS FOR SEXUAL HARASSMENT

The role of power in perpetuating sexual harassment can be viewed through two perspectives. The dominant theory of power in sexual harassment is that of the vulnerable-victim, which argues that workers in more vulnerable positions or with less power are more likely to be the targets of sexual harassment behavior (e.g., Wilson & Thompson, 2001).

Emerging from this line of thought, a common solution touted to reduce sexual harassment is then to increase the number of women in power. However, one must also consider theories of power threat, which have received considerable support (e.g., De Coster, Estes, & Mueller, 1999). The power-threat model suggests that sexual harassment may be a reaction to an individual challenging their status position. For instance, there is evidence to suggest that women in leadership positions experience more sexual harassment compared to women in non-authoritative positions (McLaughlin Uggen, & Blackstone, 2012). This may be due, in part, to a greater understanding of what constitutes sexual harassment. However, as argued by McLaughlin et al (2012), sexual harassment may also "serve as an equalizer against women in power, motivated more by control and domination than sexual desire." This was also an effect Dr. Jennifer Griffith and

continue reading on page 80 ▸▸

THIS IS AN ARTICLE BY **KELSEY MEDEIROS**

This article was first published on 16 March 2021 on *Elephant in the Lab*.

Kelsey Medeiros is an Assistant Professor of Management at University of Nebraska at Omaha where she researches workplace troublemakers. She is also the co-founder of Whisper Coalition, a research group focused on sexual harassment, and Ethics Advantage, a consulting firm working on improving workplace ethics.

I observed in our analysis of women in surgery and their experiences with sexual harassment (Medeiros & Griffith, 2019).

With this in mind, women face a double-bind with regard to power and sexual harassment. They are at risk of harassment given their typically underrepresented nature in academia, but as more women enter the academy, their very presence challenges the hierarchy, which again, puts them at risk of sexual harassment.

HOW ACADEMIA RESPONDS TO POWER MATTERS

How leaders in the academic community respond to power abuses / sexual harassment sends a signal to others as to what the culture will and will not tolerate. It is then logical that to reduce sexual harassment in the academy, we must call on our leaders to take action against this behavior. The National Science Report (2018) noted the importance of these actions in creating a culture that others perceive to be intolerant of sexual harassment and thereby reducing the behavior.

While likely an obvious solution to many readers, the reality is that leaders often fail to take action against perpetrators of sexual harassment in the workplace and specifically, the academy. One reason we often see these behaviors go unpunished is due to a perpetrator's accumulated idiosyncrasy credits (Griffith & Medeiros, 2020). Hollander (1958) coined the term idiosyncrasy credits to refer to allowances made for deviating from the norm. Specifically, credits are accumulated through good performance and high status and spent by engaging in idiosyncratic behavior. When someone with little to no credit behaves idiosyncratically, the behavior will likely be viewed as unacceptable and are likely to face consequences. In contrast, when someone who has built up a substantial amount of credit behaves idiosyncratically, they are often given a *pass*.

When it comes to sexual harassment, the use of idiosyncratic credits has been widely cited, albeit through different terms. For instance, when Susan Fowler left Uber, she noted that her claims of sexual harassment were dismissed because the harasser was a "high performer" (Scheiber & Creswell, 2017). When considering the role of power in sexual harassment, it would be remiss of us to not consider the power of the perpetrator themselves. In academia, this power most likely comes from their "star" power as researchers or teachers. Theories of idiosyncrasy credits would suggest that when these academic stars sexually harass others,

the academy is likely to dismiss accusations, "let them slide," or respond with minimal punishments.

The trio of the vulnerable-victim model, the power threat model, and idiosyncrasy credits then suggests a cyclical pattern in which women are more likely to be targeted both for their lack of, and for their possession of power, while their perpetrator may go unpunished for the very same possession of power. Academia's hierarchical nature and gender imbalance is systematically increasing women's likelihood of being harassed and ultimately, leaving the academy altogether.

WHAT WE CAN DO

It's always a soul crushing exercise to lay out just how systemic this issue is in our community. But there is certainly hope if we are willing to see the complexity of the issue and to address it head on. But this will take work and is certainly not something that will be solved by an annual 1-hour sexual harassment training. The system requires a complete overhaul.

It's also worth noting that this is not something that can be fixed by teaching women, POC, and LGBTQIA+ to avoid harassment or by "empowering" to fight back. Sure, defending and protecting yourself is great – but what if we created a space where groups of people didn't have to protect themselves from harassment on a regular basis? What if we created a space where we could all use our cognitive resources to contribute to scientific innovation and student development instead of how to avoid being harassed?

I don't know about you, but I prefer the latter. So how do we do it? What can we do? The National Academy of Sciences report notes 15 important recommendations for enacting change in our institutions. Although each are equally important and I encourage readers to review the report in full themselves, I want to raise three recommendations relevant to the preceding discussion on power.

CHANGE THE CULTURE

An organization's culture includes both tangible and intangible artifacts. With hierarchy in mind, the National Academy of Science recommends diffusing the traditional hierarchical cultural structures by creating mentoring networks and

committee-based advising. Further, shifting reward structures away from those singularly focused on publications or grants may encourage a more holistic view of faculty that spreads rewards throughout as opposed to making the academic rich richer. With regard to the tangible, it is also important for academia to examine how practices, policies, and procedures hold up the hierarchy and promote sexual harassment. When it comes to Julia's story, for instance, the practice of conducting interviews in hotel rooms created a space in which sexual harassment could more easily occur. Along these lines, recent work by Drs. Jennifer Hirsch and Shanus Khan (Sexual Citizens, 2020) on the role physical spaces play in establishing power dynamics and perpetuating sexual assault should be considered.

ACTUALLY LISTEN TO WOMEN

It is not enough to promote women. We must also listen to their viewpoints. Too often we see women placed in roles as tokens, or who are placed in roles and asked to conform to the existing norms. Women should be placed in these roles to challenge the status quo and to raise important issues that we may otherwise go unnoticed. With regard to Julia's example, for instance, men in higher status positions may fail to recognize the compromising experience of women interviewing in a hotel room. In fact, it was due to the actions of two women, Kathryn Holston and Anna Stansbury, that in 2019, the ASSA banned the use of hotel rooms for these meetings.

HOLD PEOPLE ACCOUNTABLE

In any culture change initiative, it is imperative that leaders reward the desired behavior and punish the undesirable. In this instance, leaders must not fall prey to idiosyncrasy credits, and stand up against those who engage in sexual harassment – regardless of their performance. Of course, due process and appropriate investigative procedures should be followed. Once a conclusion has been reached, however, leaders must act and signal to the community that sexual harassment will not be tolerated in the academy.

If we want to truly make academia a physically safe space for all, we need to address the root problem: power. What will you do? ◆

REFERENCES

Bacchi, C. (1993). The brick wall: Why so few women become senior academics. *Australian Universities' Review, 36*(1), 36 – 41.

De Coster, S., Estes, S. B., & Mueller, C. W. (1999). Routine activities and sexual harassment in the workplace. *Work and Occupations, 26*(1), 21 – 49. https://doi.org/10.1177/0730888499026001003

Diamond, S. J., Thomas, C. R., Desai, S., Holliday, E. B., Jagsi, R., Schmitt, C., & Enestvedt, B. K. (2016). Gender differences in publication productivity, academic rank, and career duration among US academic gastroenterology faculty. *Academic Medicine, 91*(8), 1158 – 1163. https://doi.org/10.1097/ACM.0000000000001219

Griffith. J. A., & Medeiros, K. E. (2020). Expanding the footprint of sexual harassment prevention training: A power, credit, and leadership perspective. *Industrial and Organizational Psychology, 13*(2), 137 – 141. https://doi.org/10.1017/iop.2020.26

Harned, M. S., Ormerod, A. J., Palmieri, P. A., Collinsworth, L. L., & Reed, M. (2002). Sexual assault and other types of sexual harassment by workplace personnel: A compariosn of antecedents and consequences. *Journal of Occupational Health Psychology, 7*(2), 174 – 188 https://doi.org/10.1037//1076-8998.7.2.174.

Hart, C. G. (2019). The penalties of self-reporting sexual harassment. *Gender & Society, 33*(4), 534 – 559. https://doi.org/10.1177/0891243219842147

Hegewisch, A., & O'Farrell, B. (2015). Women in the construction trades. *Institute for Women's Policy Research*. https://iwpr.org/wp-content/uploads/2020/10/C428-Women-in-Construction-Trades.pdf

Hollander, E. P. (1958). Conformity, status, and idiosyncrasy credit. *Psychological Review, 65*(2), 117 – 127. https://doi.org/10.1037/h0042501

Ilies, R., Hauserman, N., Schwochau, S., & Stibal, J. (2003). Reported incidence rates of work-related sexual harassment in the United States: Using meta-analysis to explain reported rate disparities. *Personnel Psychology, 56*(3), 607 – 631. https://doi.org/10.1111/j.1744-6570.2003.tb00752.x

Kirkner, A.C., Lorenz, K., & Mazar, L. (2020). Faculty and staff reporting and disclosure of sexual harassment in higher education. *Gender and Education*, 1–17. https://doi.org/10.1080/09540253.2020.1763923

McLaughlin, H., Uggen, C., & Blackstone, A. (2012). Sexual harassment, workplace authority and the paradox of power. *American Sociological Review, 77*(4), 625–647. https://doi.org/10.1177/0003122412451728

McLaughlin, H., Uggen, C., & Blackstone, A. (2017). The economic and career effects fo sexual harassment on working women. *Gender & Society, 31*(3), 333–358. https://doi.org/10.1177/0891243217704631

Medeiros, K. E., & Griffith, J. A. (2019). Double-edged scalpels: The trials and triumphs of women surgeons. *Narrative Inquiry in Bioethics, 9*(3), 221 – 227.10.1353/nib.2019.0057

National Academies of Sciences, Engineering, and Medicine (2018). Sexual harassment in academic science, engineering, and medicine. https://www.nationalacademies.org/our-work/sexual-harassment-in-academia

O'Connor, P. (2020). Why is it so difficult to reduce gender inequality in male-dominated higher educational organizations? A feminist institutional perspective. *Interdisciplinary Science Reviews, 45*(2), 207–228. https://doi.org/10.1080/03080188.2020.1737903

Scheiber, N., & Creswell, J. (2017). Sexual harassment cases show the ineffectiveness of going to HR. https://www.nytimes.com/2017/12/12/business/sexual-harassment-human-resources.html

Hirsch, J. S., & Khan, S. R. (2020). *Sexual citizens: A landmark study of sex, power, and assault on campus*. W. W. Norton & Company.

Wilson, F., & Thompson, P. (2001). Sexual harassment as an exercise of power. *Gender, Work, & Organization, 8*(1), 61–83. https://doi.org/10.1111/1468-0432.00122

Zhuge, Y., Kaufman, J., Simeone, D. M., Chen, H., & Velazquez, O. C. (2011). Is there still a glass ceiling for women in academic surgery? *Annals of Surgery, 253*(4), 637–643. 10.1097/SLA.0b013e318211111112

ALEXANDRA GIANNOPOULOU AND FENNIE WANG

Self-sovereign identity

Self-sovereign identity (SSI) is an identity management system created to operate independently of third-party public or private actors; it is a principle-based system backed by decentralised technological architecture, designed to prioritise security and privacy and to enhance individual autonomy.

Beyond the purely subjective sphere, identity is foundational to the status of the citizen and to the attribution of rights and duties to the citizen. Authorities attribute identities by acts of *registration*; hence, identity is inherently related to forms of socio-economic administration (The Moxy Tongue, 2012). In the digital age, the attribution and management of identity is highly dependent on private technological infrastructures. Digital identity is also an economic asset, as it consists of data that can be monetised by technology companies. First proposed by Christopher Allen in 2016, the concept of SSI emerges as a tool to emancipate individuals from the public and private interests that interact in the administration of digital identity.

Intended as a technical infrastructure for the decentralised management of user-centric, self-administered digital identities, SSI is based on ten foundational principles: 1) existence, 2) control, 3) access, 4) transparency, 5) persistence, 6) portability, 7) interoperability, 8) consent, 9) minimalisation, 10) protection. The technical dimension of SSI has so far been anchored in decentralised identifiers (DID), verifiable claims (VC) and other standards from the World Wide Web Consortium (W3C). These standards are used to link data about an identity-subject in a persistent and universal manner, allowing the identity-subject to remain in control of the information linked to their digital identity.

The set of data linked to an identity (attestations or claims) may be globally portable. Attestations may include credentials that grant the identity-subject access rights or privileges, verification of linked identity documents, professional certifications, credit history, etc. Every attestation that is linked to an identity-subject must be signed digitally by another identity-subject. The actors responsible for issuing identity elements are not stripped of their privilege, but individuals can present claims related to those identifiers "without having to go through an intermediary" (Wagner et al., 2018, p. 9). While SSI systems have traditionally been tied to the use of blockchain (in order to record transactions, sign attestations, grant or revoke access privileges), SSI is not necessarily blockchain-dependent.

There are numerous legal and socio-political shortcomings in the implementation and generalised adoption of decentralised (self-sovereign) identity. For instance, compliance with the GDPR appears to be rather challenging, due to constraints related to the governance, architecture and technological design of SSI (Renieris, 2020; Giannopoulou, 2020). Moreover, the eIDAS regulation and other sector-specific norms might demand further adaptation of SSI solutions. While SSI is

seen as a necessary solution to address the issue of undocumented populations (World Bank Group, 2018; World Economic Forum, 2018), there are considerable risks related to the expansion of SSI systems for national security and public administration. More specifically, it could be used by totalitarian governments to enforce restrictions of movement, or to identify and persecute political enemies. Moreover, the persistent integration of an internet-based identity layer cannot account for anonymity nor for the contextual, interpersonal nature of most expressions of our identity (Hopman & M'Charek, 2020).

Self-sovereign identity (SSI) is rooted in the belief that individuals have the right to an identity independent of reliance on a third-party identity provider, such as the state or any other public or private actor. Its implementation requires the development of technical standards, as well as socio-political adaptations rooted in legal amendments in order to be successful. Overall, SSI is conceived as a blockchain-adjacent but not blockchain-dependent identity management system based on user-centric design. It is intended to enable decentralised self-management of identifiers, associated with credentials and attestations that are held and controlled by users themselves.◆

The *Glossary of decentralised technosocial systems*, edited by Valeria Ferrari for Internet Policy Review, is an interdisciplinary glossary on peer-to-peer, user-centric and privacy-enhancing decentralised technologies. In order to tackle the existing gap in shared semantics, this glossary converges the efforts of experts from various disciplines to build a shared vocabulary on the social, technical, economic, political aspects of decentralised, distributed or sovereign technologies.

policyreview.info/glossary

REFERENCES

Allen, C. (2016, April 25). The path to self-sovereign identity. Life With Alacrity. https://www.lifewithalacrity.com/2016/04/the-path-to-self-soverereign-identity.html#dfref-1212

Giannopoulou, A. (2020). Data Protection Compliance Challenges for Self-sovereign Identity. In J. Prieto, A. Pinto, A. K. Das, & S. Ferretti (Eds.), *Blockchain and Applications* (pp. 91–100). Springer International Publishing. https://doi.org/10.1007/978-3-030-52535-4_10

Hopman, R., & M'Charek, A. (2020). Facing the unknown suspect: Forensic DNA phenotyping and the oscillation between the individual and the collective. *BioSocieties*, 15, 438–462. https://doi.org/10.1057/s41292-020-00190-9

Renieris, E. (2020). SSI? What we really need is full data portability. *Women in Identity*. https://womeninidentity.org/2020/03/31/data-portability

The Moxy Tongue. (2012, February 15). What is 'sovereign source authority'? *The Moxy Tongue*. https://www.moxytongue.com/2012/02/what-is-sovereign-source-authority.html

Wagner, K., Némethi, B., Renieris, E., Lang, P., Brunet, E., & Holst, E. (2018). Self-sovereign identity. A position paper on blockchain enabled identity and the road ahead. Identity Working Group of the German Blockchain Association. https://www.bundesblock.de/wp-content/uploads/2018/10/ssi-paper.pdf

World Bank Group. (2018). Identification for Development Annual Report. https://id4d.worldbank.org/sites/id4d.worldbank.org/files/2018_ID4D_Annual_Report.pdf

World Economic Forum. (2018). Identity in a digital world — A new chapter in the social contract. *World Economic Forum*. http://www3.weforum.org/docs/WEF_INSIGHT_REPORT_Digital%20Identity.pdf


#DIGITAL

THE ETHICS OF AI AND BIG DATA

A LECTURE BY JUDITH SIMON ON THE CHALLENGES FOR MORE ETHICAL ARTIFICIAL INTELLIGENCE

Advances in digital technologies in general and artificial intelligence in particular have stirred high hopes and deep fears at the same time. As a consequence, there have been calls for such technologies to be ethically designed. But how exactly can fundamental rights and moral values be accounted for in the design, development and management of systems as diverse as the Facebook newsfeed, search engines and automated decision-making tools that predict anything from creditworthiness to recidivism? In her lecture, Judith Simon explored very fundamental ethical questions about what we can do with AI for what reasons – or what we should not do with AI and for what reasons.

"Scientifically and technologically good AI is necessary but not sufficient for ethically good AI"

Judith Simon

"And what is underlying this is of course the idea that technology is not neutral or that computer systems and software are not morally neutral and that it is possible to identify tendencies in them to promote or demote particular moral norms and values."

Judith Simon

"So, what we can see here is first of all a justice problem. Societal stereotypes and prejudices, but also existing inequalities and injustices, are frequently inscribed into technologies. Intentional discrimination is possible, but it is mostly unintentional, through either the training data or different methodological choices."

Judith Simon

"Political theory and also ethics may be sources for reflection on fairness and justice and they may guide appropriate methodological choices, but these choices are always context dependent and contested. So the task of deciding on specific fairness measures should not be placed on the shoulders of developers on their own, because of their highly political character. And depending on the impact this may require very broad public debate and participation."

Judith Simon

"Ethics is in the method, but it also goes beyond it. What I mean by that is ethics within computer science education but also in the practice of designing systems is not something that comes at the end when you think about the impact. It's something you need to think about in the process of designing systems, because you need to think about it when you're choosing your data, when you're choosing your methods, when you're deciding on how to optimise and what to optimise for."

Judith Simon

"I think the notion of trustworthiness in regard to technologies only makes sense if you understand them as a socio-technical system and not as a technology per se. Because you can't trust a technology; you can rely on it, but you can only trust let's say the socio-technical network behind it, sort of like the institutions guiding it, the standards involved, the mechanisms of accountability that are behind it, but not the technology per se."

Judith Simon

"But in some instances, you may have to weigh accuracy against explainability and you may have different domains in which you can accept that stuff is not explainable if the performance is good, if the accuracy is very good. And in other domains you may decide: "Here, explainability and giving reasons for why a decision was made this way is so important that we can't rely on systems that we don't understand". This decision in itself is an ethical one – to decide "where do we need explanation?" because it's very important and "where can we give up on explanation?" because the accuracy is higher."

Judith Simon

"So there is actually a difference between explainability and accountability. You can be very much accountable for your software even if you have no clue about how this thing is working just by the sheer fact that you are deploying it. So the question of accountability must be disentangled from the question of explainability. Even if you don't explain it, you may still be held accountable. So lack of explainability cannot be a reason for not being held accountable for what you're doing."

Judith Simon

The high-profile lecture series Making Sense of the Digital Society seeks to develop a European perspective on the processes of transformation that our societies are currently undergoing. This talk by Judith Simon and all other lectures are available online.

www.hiig.de/digitalsociety

MATTHIAS C. KETTEMANN

Please prove that you are not a robot

The calls for ethics, rules and guidelines for automated decisions are getting louder and louder. But developing objective criteria for the coexistence of humans and machines is challenging. Where do we stand?

"Confirm you are human", the input screen asks before you are allowed to subscribe to a mailing list. "I'm not a robot" is what you are supposed to click. The irony grows when you consider that here, a human is confirming to a machine (strictly speaking: an algorithm-based communications application) that they are not a machine. Why, one might ask, are only humans allowed to subscribe to email lists? And what does this have to do with the fundamental questions of our society?

After the initial euphoria about the potential of artificial intelligence, the realisation has dawned that machines can get things wrong. Study after study proves it: algorithms discriminate. However, they were also developed for this purpose: to differentiate. What is important, however, are the criteria by which to differentiate, that is, the criteria according to which human and non-human decision-making mechanisms are allowed to influence each other. And what's more: neither the selection of criteria, nor the attitudes and ideologies behind the selection of criteria, are objective. And the quantities of data used to drive automated learning are rendered all the more unobjective simply because they are large. Indeed, since it is almost always historical data, it is burdened with the ideas and beliefs of the past.

GUIDELINES FOR AUTOMATED DECISIONS

The call for guidelines for algorithms, their development and use is therefore justifiably growing louder; the same applies to the right to be subject only to human decisions and to the right not to communicate with social robots (social bots) without realising it. Likewise, there are demands for the prohibition of automated decision-making systems in many social subsections – especially in the constitutional state – and for insights into the logic of automated decisions (already enshrined in current law).

In short, it seems that while, on the one hand, the digitalisation of all areas of life is continuing, with media convergence and the internet of things making our cell phones, newspapers and refrigerators targets of hacker attacks, the societal forces of inertia are gaining strength. These forces are being partly fed by actual threats, but they are also ideologically charged as a fundamental criticism of progress and technology.

We use Google to search, but Google also searches us; we use social networks, but social networks also exploit us. Drawing red lines for the technology of human-machine interaction in an ethically defensible and legally

tenable way is difficult in times of opposing approaches to prevention policy and technological progress. For this reason, we need to take a radical look at the fundamental question: how can digitalisation be designed in an ethically optimal way?

WHAT RULES DOES ARTIFICIAL INTELLIGENCE NEED?

First, ethics: ethics helps us to act properly. Ethical rules and laws shape the way we live together, help us avoid conflicts, protect rights and contribute to social cohesion. Of course, rules are subject to a constant process of change. Particularly in the field of high technology, ethics and laws must follow suit as technical applications advance. The questions we have to ask ourselves are complex: what are the criteria needed for the programming of chatbots so that they communicate fairly, that is, without unjustified discrimination? What rules need to apply when programming artificial intelligence so that it serves the good of all? How do we design the algorithms that shape our society?

Researchers at the Alexander von Humboldt Institute for Internet and Society (HIIG) are investigating these questions. The international and interdisciplinary research project "The Ethics of Digitalisation – From Principles to Practices" is a joint initiative of the Global Network of Internet and Society Research Centers (NoC). Supported by the Stiftung Mercator, they are showing how digitalisation can be made fairer. They have support from the highest level: the project is being carried out under the patronage of Federal President Frank-Walter Steinmeier, and the project's fellows have already been guests of the German head of state twice. Among other things, they showed him that digitalisation must always focus on people – something the president approved of enormously. However, the researchers at HIIG also show how research can be improved. They have shown that innovative science formats, so-called research sprints and clinics, can quickly produce vital knowledge that can then help policymakers develop better rules.

Using innovative scientific formats, the fellows were able to give the ethics of digitalisation an important update. In research sprints and research clinics, they worked out ways to legally screen algorithms for discriminatory content. They were able to identify what researchers need to know to effectively evaluate the algorithmic moderation of content on platforms (more data, especially raw data, is important; relying only on what platforms say is not enough). These findings are important and align well with the current European legislative agenda. More

continue reading on page 100 ▸▸

THIS IS AN ARTICLE BY **MATTHIAS C. KETTEMANN**

This article was first published on 20 October 2021 in the *Tagesspiegel Magazin "Berlin Science Week"*.

Matthias C. Kettemann is head of the Global Constitutionalism and the Internet research group at Alexander von Humboldt Institute for Internet and Society (HIIG) and scientific co-lead of the Ethics of Digitalisation project. He is professor of innovation, theory and philosophy of law at the Institute for Theory and Future of Law at University of Innsbruck, and research programme lead at Leibniz Institute for Media Research | Hans-Bredow-Institut.

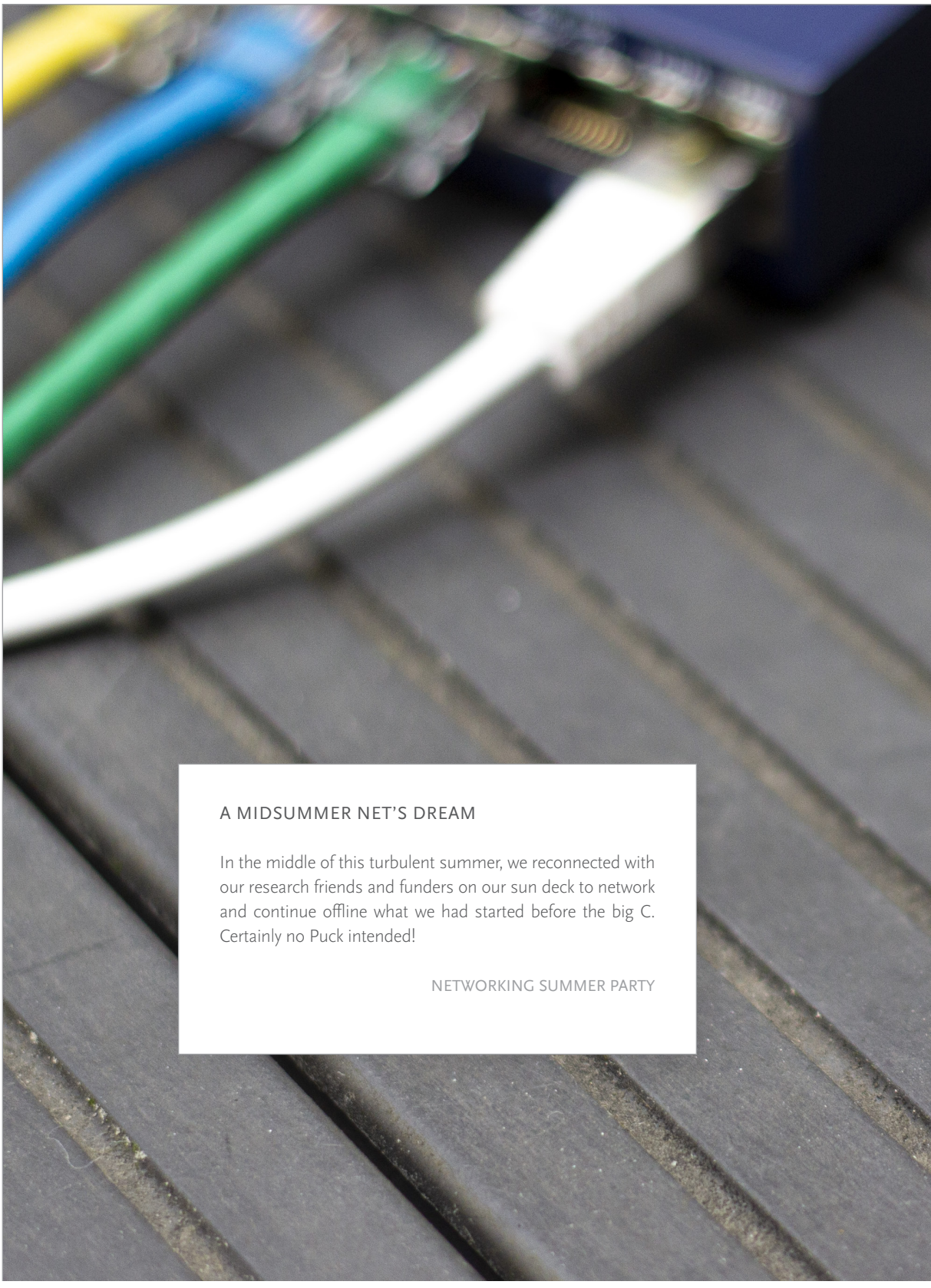
insight into the logics of using algorithms (and auditing them) is an important element of both the Digital Services Act and Digital Markets Act.

THE RIGHT TO JUSTIFICATION

In other projects, research sprints and research clinics, HIIG fellows at Harvard University's Berkman Klein Center cooperated with the Helsinki City Council to make the use of artificial intelligence in the public sector legally compliant. Incidentally, a great deal is currently happening in the area of platform research: research infrastructures are currently being established at several institutes to keep a close eye on the platforms. At HIIG, the Platform Governance Archive is making the historical terms of use of platforms searchable. At the Leibniz Institute for Media Research | Hans Bredow Institute, the Private Ordering Observatory is investigating private rule-making. Whenever an actor – be it the state or a company – has rights and obligations and assigns goods or burdens, this decision must be explainable and justifiable. All those affected by these decisions have a right to know how these decisions were made. And that is especially true when algorithms are involved.

The Frankfurt-based philosopher Rainer Forst has called this the right to justification. This right applies offline as well as online, even if the justification can be more concise depending on the situation and the decision-making processes can be faster. The project has also shown that careful governance of the way in which these justifications are communicated is always required.

What applies to algorithms was also demonstrated during the coronavirus pandemic: only when rules and decisions are understood are they perceived as legitimate. Algorithms are being used in more and more areas of society. A central task of an ethics of digitalisation is therefore to pay attention to the protection of human autonomy and dignity in such a way that our right to understand what is happening (to us) and why remains protected. ◆



A MIDSUMMER NET'S DREAM

In the middle of this turbulent summer, we reconnected with our research friends and funders on our sun deck to network and continue offline what we had started before the big C. Certainly no Puck intended!

NETWORKING SUMMER PARTY

ALI ASLAN GÜMÜŞAY AND JULIANE REINECKE

Researching for desirable futures: from real utopias to imagining alternatives

Moments of crisis may serve as critical junctures for imagining alternatives. As the future has become increasingly volatile and precarious in these unsettled times of pandemic, climate emergency, rising inequality and an ever looming digital (r)evolution, there is a great need and opportunity to develop theory that can guide society towards its future potentialities.

How can we theorise what does not (yet) exist? A central task would be to develop methodological strategies that make the future amenable to empirical study. This is quite ambitious. In this essay, we seek to take one of many steps and advocate for such (re-)search for the future, where acts of (disciplined) imagination become input for theory building.

Calls abound for us management scholars to assume a more engaged societal role by breaking away from a narrow, paradigm-driven *theory fetish* and instead contribute to solving grand challenges and societal problems (Biggart, 2016). We do not see this as an either/or. It is time for us to use the methodological and theoretical toolkit at our disposal to co-create the future and to actively feed forward soci(et)al change – not despite theory, but through it.

However, the future poses some particular problems: by definition, it is not here yet. Thus, the quest to contribute to the construction of a future social reality by theorising it raises some fundamental questions: do we actually need to wait until something exists before we can build theories about it? Or can we ex ante theorise about a post-COVID-19 world or think through the consequences of a society radically shaped by artificial intelligence? To put it differently, the conundrum we face is the following: as an empirical social science, management scholarship deals with the social world as it exists and has come to be; our methodological tools are based on data sourced from observable events that have already occurred. Thus, how can we study, conceptualise and theorise what is not (yet) observable and does not (yet) exist? Could we indeed build valid theories based on acts of imagination?

When management scholars engage with the future, their aim is commonly to anticipate possible futures through predictive analysis. But our aim is not to anticipate or predict a probable future. Instead, we seek to articulate desirable futures and how they might become reality. There are two reasons for this. First, our analytic capabilities to predict the future will likely be dwarfed by the predictive strength of corporate research. Big technology companies like Alphabet, Amazon, Facebook, Apple and Microsoft employ thousands of researchers to analyse masses of data, often routinely harvested as a by-product of machine learning. As a result, the methodological innovations needed to describe, analyse and predict human behaviour are no longer being championed by academic scholars but by capitalist institutions whose aim it is to generate profits (Savage and Burrows, 2007). Their increasingly powerful methods turn behavioural

data into what Shoshana Zuboff calls *prediction products* that not only predict our behavioural futures but also intervene in them. Outperformed by corporate research, we may find ourselves subjected to profitable but dystopian future developments.

Second, predicting the future is not (good) enough. Rather than trying to compete over who can make better predictions and build better models, we need to reclaim our societal relevance by redefining our purpose in engaging with the future altogether. A central aim must be to create more desirable futures. However, the prevailing approaches for predicting or anticipating the future in research lack such critical reflection on their normative orientation. For instance, scenario planning is a popular heuristic tool in strategy, economics and transition studies that primarily aims to generate various plausible scenarios for emerging futures. But it is precisely in the elaboration, critical reflection and theorising on futures that are not just plausible and probable but also desirable that we believe that scholarship can make a difference and reclaim its societal relevance. Rather than extrapolating to future states of the world from our present, what we also need is research that guides normative conceptions of the future. The aim would be to create new future visions – strengthened through theory – that open up radically new prospects for human agency to shape the world.

This ambition poses an obvious methodological difficulty: if the aim is to open up future potentialities that break away from the present, how can we do this using the tools of scientific analysis? The methodological challenge we hence face is to generate critical knowledge for the future with data sourced from the present. As social scientists, we commonly study the social structures of our prevailing era (Abbott, 2001). We only have empirical data when the phenomenon in question has happened. Thus, data gathering and analysis is backward looking. The predominant institutional infrastructures and settled practices that we examine also constrict, limit and even imprison our thinking and theorising. But we want to look forward. And to do so, we need to free ourselves from our own cognitive and methodological chains. Can we do so while maintaining standards of academic rigour?

FROM REAL UTOPIAS TO ACTS OF IMAGINATION

One response is to study real utopias. Real utopias exist on the fault line between "dreams and practice" (Wright, 2010, p. 3). They are utopian because they involve

developing visions of future alternatives to predominant institutions. But they are also real because they are rooted in the potentialities of the present. To study real utopias, many of us have focused on alternative forms of organising, such as spiritual, ecological and social collectives, communities and cooperatives that exist on the periphery of the mainstream. For instance, in our own work, we studied a digital social incubator (Gümüşay and Smets, 2020), an Islamic bank with a values-based business model (Gümüşay et al., 2020), prefigurative organising in the Occupy movement (Reinecke, 2018) or fairer trade (Reinecke and Ansari, 2015). These examples demonstrate, on a small scale, what could be possible. Yet, the main practical and conceptual challenge with these alternative forms of organising is precisely that they are *alternatives*; they take the form of small-scale social enclaves on the periphery. They need to be *translated* for the centre of society, or scaled up, without losing the essence of what rendered them inspiring visions of the future in the first place. Still, focusing on concrete examples of real utopias allows us to access existing empirical data while also generating novel insights into the possibilities of creating more sustainable or equitable future organizations. We as academics can then co-create social change towards a desirable future by theorising and legitimising its occurrence on the fringes.

We believe that there is another way of researching the future that goes beyond the search for existing empirical alternatives: researchers can feed forward soci(et)al change through acts of imagination about the future. Imagination refers to the ability to form pictures of something that cannot be immediately sensed or that has not been previously perceived: the irreal, unreal and surreal. Imagining is making the absent present. However, the validity of theories based on imagination cannot be tested against the empirical present and might be deemed pseudoscience. Such acts of imagination therefore require a new methodological toolkit to achieve speculative rigour. As an academic approach, we need disciplined imagination that not only considers what is feasible and probable but also what is desirable.

Acts of imagination can be radical because they move away from a reliance on empirical data about the present and venture into the domain of imagination. Such research must logically be based on forms of fictional empirical data. It may be found in places of *forethought*, including calls for action and manifestos, such as *democratising work* or *decolonising the university*. These are, by definition, still fictional. If we do not wish to wait until something exists in order to theorise it, then how can we build valid theories based on disciplined imagination while maintaining scholarly rigour? How we respond to this question has important

methodological implications for the possibilities of empirical inquiry and the purpose of theory building – to which we turn next.

(RE-)SEARCHING FOR THE FUTURE

As a starting point for developing methodological strategies that make acts of imagination amenable to empirical study, we see two pathways. The first pathway would start from alternatives that already exist at the fringes of the mainstream – real utopias – and imagine what impact these alternatives would have if they had broader or even universal reach. In other words, we need to examine what would happen if utopian social enclaves scaled up and became reality at a broader level. For instance, some organisations already organise themselves in a circular fashion, but how would an entire economy be organised to achieve circularity? Or how could the widespread use of self-driving cars transform mobility and improve – or worsen – human behaviour?

The second pathway would involve seeking ways to explore imagined alternatives that do not (yet) exist. For instance, how would our modes of organising be impacted if artificial intelligence came closer to the threshold of singularity? What would further gene manipulation or advancements in robotics mean for organisations and work? By exploring developments before they are reality, these mind-made imaginaries could open up possibilities, inspire people and orient action. Thus, they have a pre-prefigurative potential; they allow us to imagine the enactment of visions of the future and also *backcast* socio-political practices that would permit us to prefigure such a future.

To begin this process of imagining, we could engage with the central grand challenges of our time by studying social movements such as Fridays for Future or Black Lives Matter in greater detail. Research could imagine their demands being materialised and think about how this would impact society. Such research would require future-perfect thinking – thinking backwards from a possible future. It could entail imagining an ecotopia where all resources are renewable, sourced from cradle to cradle in a circular economy, and then ask: what does it mean for individuals, organisations and society – and our theorisation these? It would not only entail analysing Martin Luther King's speech but also exploring what would happen if his dreams of a racism-free world became true. Such narratives constitute aspirations for a better world through ideas whose time have not (yet)

continue reading on page 110 ▸▸

THIS IS AN ARTICLE BY **ALI ASLAN GÜMÜŞAY AND JULIANE REINECKE**

This article was first published on 28 March 2021 in the *Journal of Management Studies*.

Ali Aslan Gümüşay is head of the Innovation, Entrepreneurship & Society research group at Alexander von Humboldt Institute for Internet and Society (HIIG) and a postdoctoral researcher at University of Hamburg. His research consists of four streams: values, meaning and hybridity in entrepreneurial settings; grand challenges, innovation and new forms of organising; societal complexity and engaged scholarship; and digitalisation, AI and the future of work.

Juliane Reinecke is professor of international management and sustainability at King's College London. Juliane's research draws on insights from organisation theory, political philosophy and process studies to explore, broadly speaking, how transnational governance institutions emerge and evolve as a result of the interactions of multiple stakeholders to promote more just and sustainable forms of globalisation in global supply chains, but also in organisations in general.

come. Many describe states of human flourishing and well-being, socio-political equality, human-nature symbiosis and human-centred technological progress.

We could also focus on new *research sites* where acts of imagination would take place and where actors would engage in projective deliberation in community forums, social movements, citizen dialogues or policy arenas with the aim of imagining and elaborating possible futures. We could expand our methodological toolkit using future-oriented living labs, that is, *future labs* that act as spaces for the creation of thought experiments or utopian thinking. The advantage of studying such sites is that, in their externalisation in actors' talk, text, expressions and narratives, they make imagined futures visible and empirically accessible. Analysis could focus on the fictional stories, scenarios, or maps for action that are constructed conversationally. Of course, we would need to be aware that imagination is contingent on the dynamics of interaction and experiment with settings that encourage acts of imagination. In contrast to conventional methods, such as Delphi studies, which try to serve as oracles that forecast and predict future realities as accurately as possible, acts of imagination do not serve to predict probable futures but to articulate desirable futures and then, backcast as to how they might become more likely. While multiple methods aim to forecast a future and examine feasibility and probability, we see a need for acts of imagination, in particular related to desirability.

FROM POST-FACTUAL TO PRE-FACTUAL

We hope that this essay will open up the conversation about what a new future-oriented research agenda might look like, and we hope that our colleagues will join us so that, together, we can develop new ways to research (for) the future. It is clear that this must engage the community of management scholars collectively, because the implications of re-imagining what data is, how it is used and how we theorise are far reaching. This re-imagining would offer a complementary way of doing research in which imagination becomes data and where we see mind-before-matter. While we currently face the challenges of a so-called post-truth world, we may rigorously work on a pre-truth world: from post-factual to pre-factual.

To conclude, we argue that we need to imagine alternative futures and that this requires us to renew our methodological toolkit and rethink the purpose of theorising in terms of performing desirable futures. Our academic profession is uniquely placed to do so because of our distinct ethos – we are not driven by

profit objectives but rather by scientific and societal norms. We thus envision that adopting both a critical and normative stance will become more important. This will require an awareness of alternatives and conscience in engaging with them. If we don't imagine the future, others like technology companies will. We need to think more about our role as an intellectual conscience that bridges head and heart. Academia is a vocation, a profession to be professed. Otherwise, we will be outsmarted – only to study, explain and theorise social realities that were imposed on us. To this end, we propose disciplined imagination of alternative, desirable futures as a form of avant-garde research that does not only examine reality-in-the-making but can also shape social reality through its performative potential. Imagine that. ◆

REFERENCES

Abbott, A. (2001). *Time Matters on Theory and Method*. University of Chicago Press.

Biggart, N. W. (2016). 'Biggart's lament, or getting out of the theory cave'. *Journal of Management Studies*, 53, 1381–87.

Gümüşay, A. A., & Smets, M. (2020). Ney Hybrid Forms and Their Liability of Novelty. *Research in the Sociology of Organizations*. 69: 167–187.

Gümüşay, A. A., Smets, M., & Morris, T. (2020). "God at Work": Engaging Central and Incompatible Institutional Logics through Elastic Hybridity. *Academy of Management Journal, 63*(1): 124–154.

Reinecke, J. (2018). Social Movements and Prefigurative Organizing: Confronting Entrenched Inequalities in Occupy London. *Organization Studies, 39*(9): 1299–1321.

Reinecke, J., & Ansari, S. (2015). What is a "Fair" Price? Ethics as Sensemaking. *Organization Science, 26*(3): 867–88.

Savage, M. and Burrows, R. (2007). 'The coming crisis of empirical sociology'. *Sociology*, 41, 885–99.

Wright, E. O. (2010). *Envisioning Real Utopias*. Verso.

HIIG × ANNIVERSARY

ON THE LEAP INTO THE FUTURE

BURNING ISSUES FOR THE NEXT TEN YEARS OF INTERNET AND SOCIETY RESEARCH

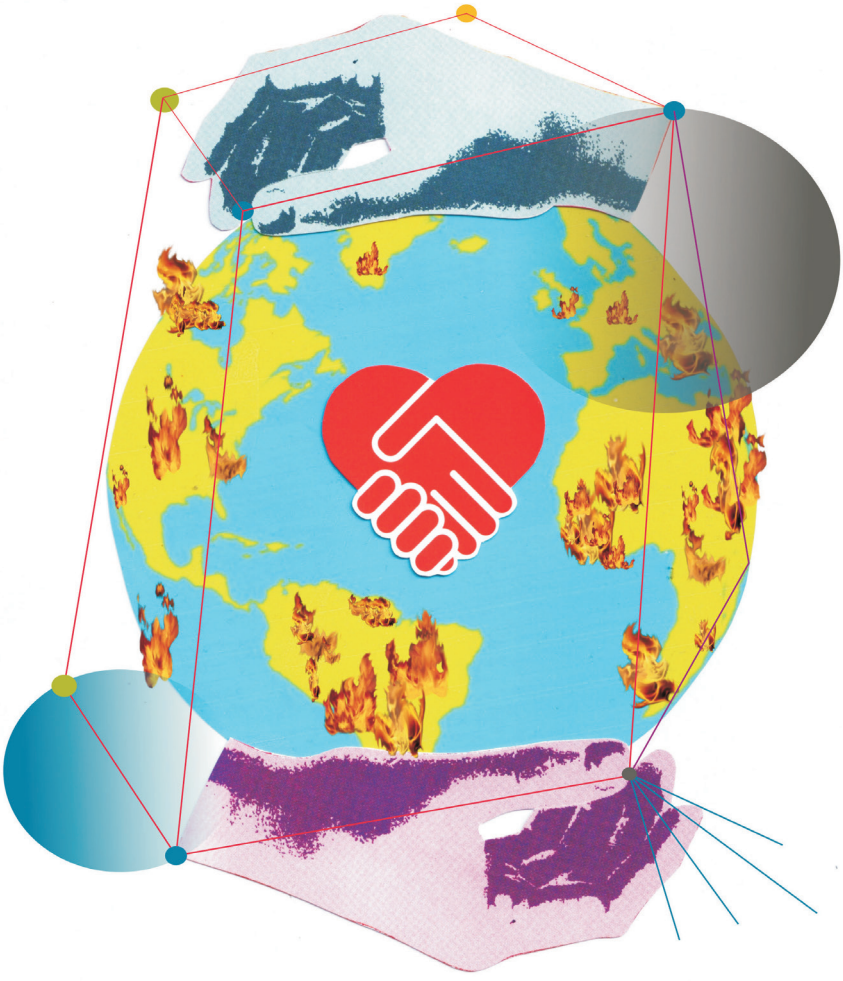
The digital transformation of our society is far from over and digital innovations and internet technologies will likely be able to make a real leap forward again in just a few years. Looking at the changing conditions and challenges of our interconnected world, what topics will we be dealing with in the future at the Alexander von Humboldt Institute for Internet and Society (HIIG)? What questions need to be answered? On the occasion of HIIG's 10th anniversary, we reflected on the topics that will most likely occupy our digital society for the next ten years. To launch the HIIG X Anniversary series, our directors Jeanette Hofmann, Thomas Schildhauer and Wolfgang Schulz give insights into key themes and how they relate to internet and society research.

POWER

Technologies are about power. Who gets what, when and how? And who is excluded? With digital platforms acting as proprietary markets and powerful big tech, we need to question who should provide the essential infrastructures that form part of the fabric of our societies.

"The data economy has brought about new forms of market power and its abuse. To a large extent, the abuse of market power rests on information asymmetries resulting from the big data amassed by a handful of global platforms. Data have become an essential resource for commodifying information about people but also for training the machine learning systems that have been designed to predict future trends and automate decision-making. However, data monopolies do not only disadvantage individual users and businesses but also societies at large. Platforms tend to analyse data for rather narrow business-related purposes, thereby leaving their data troves systematically underexploited. There are so many questions societies could ask, questions that would offer them a better understanding and inform better actions, if only they had access to this remarkable resource. It is about time to strengthen the regulatory counter power to release the potential for prosperity that is locked up in data monopolies."

Jeanette Hofmann



HUMAN RIGHTS

Human rights are influenced by the internet on a large scale. Take the so-called Arab Spring, election meddling or clickwork – unrest and violations of rights are not only the product of real-life occurrences. They are closely linked to the world wide web. Digitalisation is a double-edged sword that can both create accountability and control mechanisms, which can be (mis-)used by authorities.

"Human rights are rightly regarded as a great civilisational advance of mankind. While, in principle, they have great significance in the intergovernmental sphere, this is practically limited in many fields. Due to conflicts on global communication platforms, they have again become a more frequent topic of discussion. The question of the platforms' commitment to human rights and what this means in concrete terms becomes central to communications law. In order to give traction to human rights, there is a need for institutional structures that ensure their application. Here, there is a field of research that is just opening up and will have to be worked on in the coming years. It is interesting to note that the reference to human rights can be found in state regulatory arrangements, but also in the private law-making of platforms. An area of hybrid governance – as we call it – is emerging in which human rights play a central structuring role."

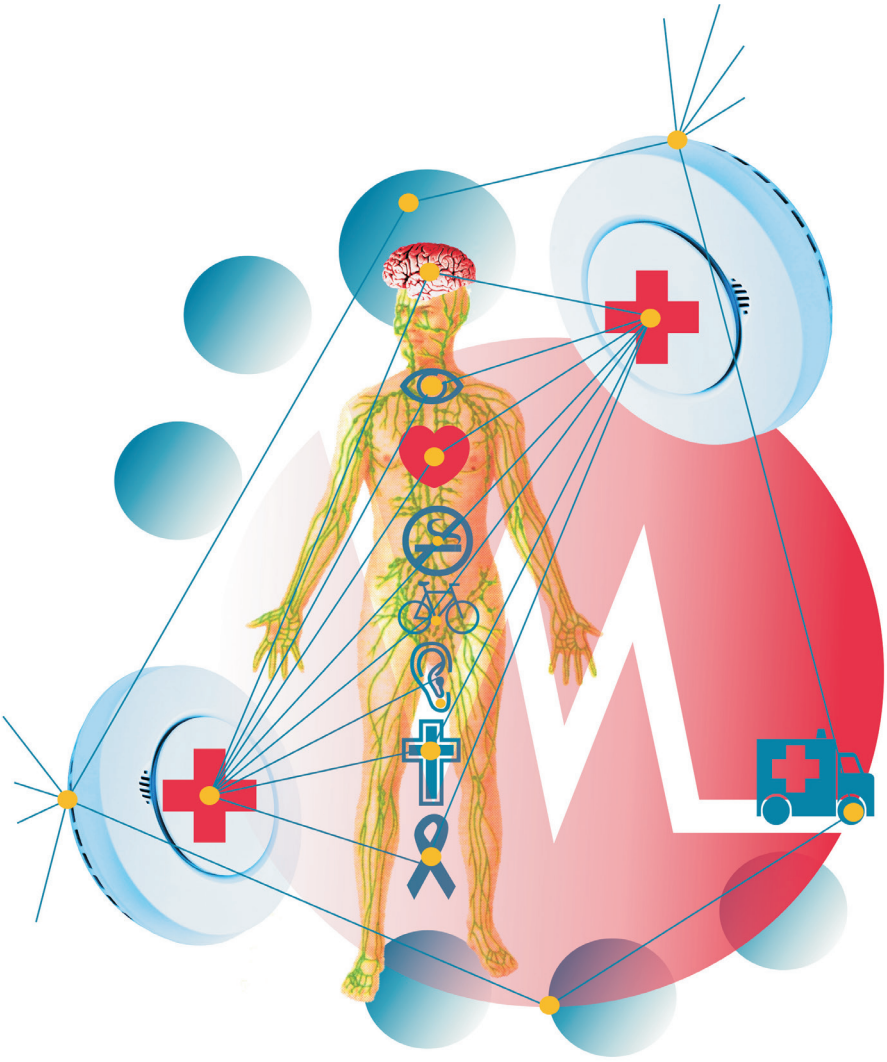
Wolfgang Schulz

HEALTH

With new digital technologies, we are also approaching novel ways to shape our life in terms of health, aging and well-being. What is desirable, how do we want to implement it and do we also want to set limits on what is possible? Such questions need more attention.

"We see a world in which people live longer, but what can digital technologies contribute to also help people to live better and how could AI, big data, platforms, robotics or blockchain make life in an aging society easier. The questions are wide-ranging: how can older people age on their own terms; and, more generally, how can changing demographics be addressed? We not only argue that those topics are highly relevant but also seek to support digital and internet-based innovations in the healthcare sector and provide answers to questions that currently go unaddressed due to limited research, data security issues and unclear user needs. For this purpose, the HIIG-initiative Digital Urban Center for Aging & Health (DUCAH) has recently built a unique structure which, through an open innovation method, has brought together the wishes and needs of people and our interdisciplinary research expertise. Together with start-ups, social economy organisations, health-care companies and other research institutions, we seek to analyse, develop and support people-centred solutions to our digital life in terms of health, aging and well-being."

Thomas Schildhauer



BRONWEN DEACON AND MORITZ TIMM

Possibilities for change: higher education and digitalisation

As the first digital semesters come to an end, the authors take a look at where higher education institutions stand, what changes have been made to the system and how the system needs to be further reformed if we want to successfully bring higher education into the digital age.

The education sector recently found itself in troubled waters: the pandemic revealed – and in some cases is still revealing – enormous delays in the digitalisation of German higher education institutions (HEIs). The abrupt, pandemic-related closures of HEIs has left much of the German higher education sector overwhelmed, which seems partly due to their lack of expertise in online learning and their sluggish moves towards new modes of teaching. This slow pace of digitalisation has manifested itself mainly in inadequate infrastructure (Gillmann, 2017) and the absence of a culture of digital innovation in HEIs (Bils et al., 2019).

AN OLD AND PROVEN SYSTEM IS FORCED TO CHANGE

German universities have a rather good reputation: out of the best 200 universities in the world, a significant number are located in Germany (universityrankings.ch, n.d.; timeshigherducation.com, n.d.; roundranking.com, n.d.). University education in Germany is inexpensive, largely because around three quarters of Germany's 400 or so universities are state run, and studying there means paying only small administrative fees of a few hundred euros per semester.

This system was subjected to a massive shock due to the pandemic. Even though digital teaching is not new in German higher education, HEIs, especially the state-run variety, have seemingly done little to foster a culture of digital teaching and learning. Before COVID, lecturers had few educational opportunities to learn the necessary digital skills and expertise. In addition, there is a need for digital applications such as forums, online exams and assessment options that could challenge traditional understandings of education by focusing more on the user-oriented perspective (Bils et al., 2019). This is despite the potential of digital teaching, promoted not least by educational technology (EdTech) providers. This potential first and foremost concerns flexible access to education – which has now become crucial since the closure of universities.

The scholarly debates on the subject were heated, but the fact remains: EdTech was prominent and visible for years prior to the pandemic (Laufer et al., 2021). Brick-and-mortar universities, however, did not engage with these possibilities and therefore did not adopt EdTech more broadly. A 2018 strategy paper by Hochschulforum Digitalisierung, a think tank concerned with digitalisation strategies for HEIs, stated that it would take time to establish a clear strategy and engage with committed people on all levels of

HEIs: "German HEIs largely have internal motivations for engaging with digital teaching. There is almost no external pressure to redesign teaching in the digital age" (Schünemann & Budde, 2018, p. 11). Since this almost prophetic statement was made, one thing has changed: COVID-19 is probably the biggest external pressure German HEIs have felt in decades.

EMERGENCY REMOTE TEACHING VS. ONLINE LEARNING: A QUESTION OF PREPARATION

What appeared to be an old but robust system pre-COVID suddenly seemed problematic: HEIs needed a response to the closure of their institutions – a quick way into digital teaching. As we previously noted (Deacon & Timm, 2021), most HEIs adopted crisis management approaches – which are by nature quick, ad-hoc and tailored to a specific case. Those involved in crisis management do not seek to change the status quo but rather to maintain it in times of sudden change. We might hence describe what German HEIs did as emergency remote teaching (ERT), a concept usually applied to armed conflicts, natural catastrophes or political upheaval (Schlesselmann, 2018). In ERT, teaching is quickly migrated to the digital realm without the option to set time and resources aside. In such circumstances, teaching continues, but learners' needs are neglected, largely because proper online learning usually is not just a digital mirror of regular, analogue teaching. The fundamental question at this point is: do German HEIs want to overcome their crisis-response mode or do they want to change fundamentally?

German HEIs struggled with their crisis responses to a point where a large group of acclaimed professors called for a *non-semester* in an open letter last year, which was backed by the free union of student bodies. The initiative demanded that the summer semester of 2020 not be counted for degree-awarding purposes. The reasoning behind it was rather simple: the authors believed that e-learning was just not enough of a thing in Germany. The claim that "most learners and teachers are not sufficiently used to the methods of online teaching" was a core point in their argument for striking the semester. This did not come as a surprise, considering how little German universities had dabbled in EdTech and digitalisation. Yet, despite their crisis-response mentality, German HEIs were willing to experiment with new formats, platforms and tools. Logically speaking, an intermediate step was required in the transition between analogue and digital teaching – in German HEIs, that step was ERT. However, the question remains

continue reading on page 124 ▸▸

THIS IS AN ARTICLE BY **BRONWEN DEACON AND MORITZ TIMM**

This article was first published on 19 July 2021 on *Elephant in the Lab*.

Bronwen Deacon is a doctoral student and researcher within the Knowledge and Society research programme at Alexander von Humboldt Institute for Internet and Society (HIIG). Her research focuses on the adaptability of higher education institutions in times of digitalisation and the involvement of different stakeholders and especially different types of leadership in such change processes.

Moritz Timm is a researcher within the Knowledge and Society research programme at HIIG. His research interests include the digital transformation of the higher education sector and algorithmic injustices and scepticism towards artificial intelligence.

whether these new modes were meant to be long-lasting. The fact alone that the semester was not struck out for degree-awarding purposes suggests that what was done spontaneously worked out to a certain extent.

The counter-concept to ERT would be online learning, a method which is, just like regular face-to-face teaching, centred on a sustained transmission of knowledge. Here, as mentioned above, digital curricula are carefully crafted, making the most of EdTech's features while also assessing learning goals and measuring student success. This indeed requires professional training, a detailed planning phase and, of course, evaluation. Such a system requires a long design period and is nearly impossible to set up ad-hoc. Online learning therefore focuses on the learning rather than the teaching part of the student-teacher exchange and aims to achieve high quality learning outcomes (Fenstermacher & Richardson, 2005) rather than just doing the teaching in a digital manner.

In contrast to what we have described in public HEIs, the private tertiary education sector seemed less affected or perhaps better prepared. Roughly three quarters of private HEI leaders stated that the pandemic has not negatively affected their institutions, according to Stifterverband (2021). 97.6% of them believed that switching to online learning was going smoothly. Private HEIs had more often adopted concepts like online learning before the crisis. Most of them had been offering asynchronous learning opportunities for years, developing an approach to teaching which seems largely foreign to state-run universities (Sperlich, 2008, p. 132). Moreover it is rather obvious that, due to the market-driven nature of the private education sector, their need to adapt was more intense – after all, students are paying significant amounts, forcing their HEIs to provide modern teaching methods more quickly than public sector institutions – in this case, too, external pressure drove change processes.

WAKING UP: COVID-19 AND ORGANISED FREEDOM AS AN OPPORTUNITY

As discussed above, responses to the pandemic differed from case to case. For example, not all academic disciplines are in a position to respond in the same way: fields like medicine, the arts or some natural sciences need face-to-face lessons or access to laboratories and equipment, which can make online learning more difficult. All in all, we can see two patterns of how HEI stakeholders look at the future of (digital) teaching: There are those few institutions that are merely

passing the time until HEIs reopen to return to the old normal – resistance to change at HEIs is well known (Anderson, 2008; Krücken, 2003). The majority, though, seem to see opportunities to adapt their teaching – for example, by expanding their use of hybrid courses and flipped classroom models, thereby changing their established routines. To make this work, organisational changes in HEIs have to enable the expansion of online learning.

In a previous study conducted within our Organisational Adaptivity in Higher Education Context (OrA) research project, we formulated a concept called organised freedom based on the individual needs of teachers in HEIs and the relevant organisational change processes (Elsholz et al., 2021). In the study, we set out to understand what enables the sustainable digital development of teaching. We found that HEIs need to provide a framework consisting of infrastructure, resources and a supportive culture. Additionally, the time and freedom to redesign teaching needs to be allocated. Organised freedom can be viewed as one of many under-researched success factors necessary to achieve long-term organisational flexibility, thereby paving the way for digital development within the higher education sector in Germany and elsewhere. While the literature discusses HEI management and students (Kebritchi et al., 2017), we focused on the actual practicalities of teaching online.

If we are to attempt a shift in higher education towards new modes of learning and teaching, lecturers and institutional management will first need to adapt. While management can, as stated above, give teachers the freedom to plan, teachers themselves need to change too. For them, a shift in how they view themselves and their tasks is necessary. They need to reconceptualise their own roles – teaching, in the analogue world, used to be a performative act while now, speaking digitally, teachers will need to take the role of a guide rather than performer (Kebritchi et al., 2017). Additionally, HEI managers will have to provide a framework that is conducive to innovation in teaching and learning. With the power to open and close communication channels and the ability to define university life, such a framework could pave the way for a new era of education by enabling it to adapt to current challenges (Kiendl-Wendner, 2016).

So teachers need to reassess their role and HEI managers need to do so too. As long as this area remains a battleground for progressives and conservatives, the opportunities for change seem low. ◆

REFERENCES

Anderson, G. (2008). Mapping academic resistance in the managerial university. *Organization, 15*(2), 251–270.

Bils, A., Brand, H, Pellert, A. (2019). Hochschule(n) im digitalen Wandel. Bedarf und Strategien. *Aus Politik und Zeitgeschichte 27–28*(2019). https://www.bpb.de/apuz/293132/hochschulen-im-digitalen-wandel

Deacon, B., & Timm, M. (2021). Possibilities for Change - Higher Education and Digitalisation. *Elephant in the Lab*. https://doi.org/10.5281/zenodo.5113453

Elsholz, U., Fecher, B., Deacon, B., Schäfer, L. O., Laufer, M. (in press). Organisierte Freiheit!? Implikationen der COVID-19-Pandemie für digitale Lehre: Organisierte Freiheit als Veränderungsparadigma. *MedienPädagogik*.

Fenstermacher, G., Richardson, V. (2000). On Making Determinations of Quality in Teaching. *Teachers College Record 107*(2005). 186–213.

Gillmann, B. (2017, July 24). Der digitale Nachzügler. *Handelsblatt*. https://www.handelsblatt.com/politik/deutschland/deutschland-und-die-digitalisierung-der-digitale-nachzuegler-/20098982.html

Kebritchi, M., Lipschuetz, A., Santiague, L. (2017). Issues and Challenges for Teaching Successful Online Courses in Higher Education: A Literature Review. *Journal of Educational Technology Systems, 46*(1), 4–29.

Kiendl-Wendner, D. (2016). Die Qualität der Hochschullehre und deren Messung. In Steirische Hochschulkonferenz (Eds.), *Qualität in Studium und Lehre* (pp. 243–261). Springer.

Krücken, G. (2003). Learning the 'New, New Thing': On the role of path dependency in university structures. *Higher Education, 46*(3), 315–339.

Laufer, M., Leiser, A., Deacon, B., Perrin de Brichambaut, P., Fecher, B., Kobsoda, C., Hesse, F. (in press). Digital Higher Education: A Divider or Bridge Builder? Leadership perspectives on EdTech in a COVID-19 reality. *International Journal of Educational Technology in Higher Education*. https://doi.org/10.1186/s41239-021-00287-6

Marczuk, A., Lörz, M. (2021). Studieren unter Corona-Bedingungen: Studierende bewerten das erste Digitalsemester. *DZHW-Brief, 05*(2020). 1–7.

Schlesselmann, S. (2020). Perspective from a Teaching and Learning Center During Emergency Remote Teaching. *American Journal of Pharmaceutical Education 84*(8). 1042–1044.

Schünemann, I., Budde, J. (2018). Hochschulstrategien für die Lehre im digitalen Zeitalter. (Hochschulforum Digitalisierung Arbeitspapier Nr. 38). https://hochschulforumdigitalisierung.de/sites/default/files/dateien/HFD_AP_Nr38_Empfehlungen_Strategieentwicklung.pdf

Sperlich, A. (2008). Private Hochschulen in der deutschen Hochschulforschung. Oder: "Noch nicht mal ignorieren". *Die Hochschule: Journal für Wissenschaft und Bildung 17*(2008). 126–139.

Stifterverband (2021). Der private Hochschulsektor in der Pandemie. Derzeitige Auswirkungen und Perspektiven. *Stifterverband Policy Paper 02*(2021). 1–9.

SONJA KÖHNE AND MIRIAM KLÖPPER

Working from home but never alone

People analytics have the potential to support and empower employees working from home. To prevent discrimination against employees or invasions of their privacy, clear regulation combined with employee participation is needed. Therefore, both works councils and employees themselves need more thorough information about the consequences of data collection by digital means and about their own rights in this process.

The COVID-19 pandemic has left us with many changes in what we called our everyday lives. One of the most significant changes concerns how we organise our work. At the end of March 2020, about one quarter of employees in Germany were working from home (Möhring et al., 2020). With this, remote work lost its reputation as being a lifestyle choice and quickly became a political issue. One major concern in the public debate on the right to work from home was the working conditions people face when they are physically separated from their team. With communication taking place solely by digital means and without traditional supervision, both employers and employees had to rethink their work routines. People analytics tools offer a data-driven approach to human resource management (HRM) and promise to optimise employee-related business decisions and to facilitate leadership – also from a distance. They collect data generated in employees' day-to-day work – for example by tracking the number of emails, calls or meetings a person participated in. Based on this data, the tools use algorithms to analyse employees' productivity or potential, make recommendations on how to improve performance (e.g. indicating productivity peaks and lows during the day) and in some cases even predict who will likely be a high/low performer in the future. The data often stem from popular applications like Slack or Microsoft Office. Other vendors like Humanyze analyse communication patterns and relationships within the company: which teams like to collaborate? Where are knowledge silos emerging? However, these tools in themselves are no cure-all for the loss of face time. They need to be designed and employed carefully and deliberately. Hence, driven by the pandemic, a new need for action has arisen and with it the necessity for a thorough examination of the potential and challenges of remote work.

CHALLENGES OF REMOTE WORKING

Taskin and Devos (2005) have identified three specific tensions in the despatialisation of remote work. First, there is an intensification of work, where tensions arise between professional and private time. Thus, the reduction of work-related stress by opting for remote work appears to increase stress levels in private life. Second, there is social isolation, which – despite high levels of autonomy in remote work – has a negative impact on commitment and identification with the company and can thus negatively influence work performance. Third, there is what they call the *do-it-yourself rule*, which describes extensive control and monitoring methods. These stand

in stark contrast to the frequently made claims that employees can self-manage in remote work. Challenges in supervising employees remotely and the field of tension between autonomy and control strongly dominate the existing research on remote work.

QUESTIONS OF POWER AND CONTROL

In the 2016/17 Linked Personnel Panel in Germany, two-thirds of employees who never worked from home cited their supervisors' desire for them to be present as a reason for not working remotely (Grunau, Ruf, Steffes & Wolter, 2019). Similarly, one in ten companies that did not offer remote work cited difficulties in management and control as a reason for not offering it. Thus, the spatial separation of supervisors and employees in remote work goes hand in hand with questions of power and control. Supervisors face a major loss of control when employees suddenly move outside their spatial reach. Information asymmetry in remote work is high (Cristea & Leonardi, 2019), making it difficult to assess whether employees or colleagues are particularly hard-working or committed at any given moment (Leonardi, Treem & Jackson, 2010). Due to the physical separation, the work outcomes are one of the few ways for supervisors to appraise their employees. The increased autonomy of employees vis-à-vis their superiors in remote work diminishes the role of supervisors in regard to their employees' professional concerns.

TURNING TO TECHNOLOGY

To compensate for the loss of control, employers can turn to technological tools. Although they are currently not prevalent in Germany, these technologies are becoming increasingly powerful. Yet, there has been a lack of in-depth theoretical understanding of how they affect leadership dynamics (Avolio, Sosik, Kahai & Baker, 2014). While the use of communication technologies gives employees in remote work a sense of increased autonomy, it also creates new constraints in an environment where employees could previously escape the control of their employer (Sewell & Taskin, 2015). Involved actors understand these digital technologies as social rather than purely technical and thus as a manifestation of a "society of control" (Martinez, 2011). In this sense, the newly gained spatial flexibility of remote work comes at a price: employees must henceforth navigate various forms of control. Moreover, these technologies enable supervisors to invade domestic or private spaces (Fleming & Spicer, 2004). Remote work

thereby changes the management function in terms of roles, expectations and relationships. Yet, organisational control certainly does not disappear, but merely becomes "more implicit, political, social, even cultural" (Taskin & Devos, 2005, p. 20). Aside from the increased use of communication technology, there is also a growing debate around people analytics tools that promise to make up for the loss of control during remote work.

PEOPLE ANALYTICS IN A NUTSHELL

Tursunbayeva et al. (2018) define people analytics as "an area of HRM practice, research and innovation concerned with the use of information technologies, descriptive and predictive data analytics and visualisation tools for generating actionable insights about workforce dynamics, human capital, and individual and team performance that can be used strategically to optimise organisational effectiveness, efficiency and outcomes, and improve employee experience" (p. 231). Thus, people analytics tools are designed to measure and analyse employee-generated data. For office workers, the data is usually collected through software applications, although tools using wearables to track warehouse workers also exist (Jarrahi et al., 2021). There is a full range of tools available today that capture and analyse variables like the duration and use of individual applications, email traffic or the number of meetings held and combine it with personal data such as age, gender or the distance between home and workplace. These applications aim to support all HR core tasks with a data-driven approach by analysing employee potential, identifying training needs or predicting fluctuation. They thereby inform core managerial decisions, such as whom to hire, promote or fire. One popular use case pertains to recruiting: people analytics tools scan applicants' CVs, compare them with existing data about high performers within the organisation and make hiring recommendations. The tools have, however, been the subject of a wide public debate. One of the most widely known people analytics software solutions, Microsoft Workplace Analytics, made headlines at the end of 2020 for incorporating a productivity score in their software. Microsoft quickly reacted to public criticism and no longer allows employers to see individual employees' scores. In Germany, the case of Zalando's people analytics tool Zonar raised concerns: employees of the fashion retailer were evaluated and ranked by their colleagues with it – just like products in an online shop. This led to an increased level of stress among workers (Staab & Geschke, 2020).

MANAGING FROM A DISTANCE WITH PEOPLE ANALYTICS

As mentioned above, remote work creates a new demand for technologies that enable companies to supervise and control employees in dispersed teams. People analytics tools therefore seem to be an obvious choice to track the productivity of employees who are working from home. For example, they enable supervisors to see if a team has worked their scheduled hours and kept up productivity. Yet, it is not just supervisors who might want to see the hours a team has worked – works councils, too, can use this data to assess if employees take sufficient breaks and prevent them from working too many hours. For employees, the introduction of remote work may lead to higher expectations regarding their availability (BMFSFJ, 2017) and thus to increased pressure to perform and consistently work overtime. People analytics may further be adapted to assess employees' well-being and stress levels, thereby generating early warnings regarding potential risks of burn-out (Banholzer, Feuerriegel, Fleisch, Bauer, & Kowatsch, 2021; Estévez-Mujica & Quintane, 2018). These features appear to be particularly beneficial during remote work, where in-person communication is rare. When working together on-site, a supervisor is able to gather a majority of this information by talking to their staff or seeing them performing tasks at their desks. People analytics can thus become an enabler of remote work, by supporting self-organisation and supervision. However, it can also pose a threat to employees.

THE DARK SIDE OF PEOPLE ANALYTICS

Although there are grounds for expecting benefits of the use of people analytics for teams working remotely, there are various reasons that indicate the contrary. People analytics enable employers and supervisors to gain insights into various domains that have previously been private. Thus, there is a strong asymmetry regarding the information available to employees and employers; in some cases, employees cannot even verify what kind of data and information is gathered about them. Another valid argument against the use of people analytics is that some of the gathered data need not be made available to managers. For example, a software can simply remind people to take breaks instead of alerting the employer, who may not even act on this information. Measuring productivity by the amount of phone calls and emails, as some people analytics tools currently do, often does not provide accurate information about a person's actual productivity. On the contrary, this approach can cause employees to make unnecessary calls and write pointless emails because of the pressure to appear productive. In this way,

continue reading on page 134 ▸▸

THIS IS AN ARTICLE BY **SONJA KÖHNE AND MIRIAM KLÖPPER**

This article was first published on 12 April 2021 on the *Digital Society Blog* of the Alexander von Humboldt Institute for Internet and Society (HIIG).

Sonja Köhne is an associated doctoral researcher in the research group Innovation, Entrepreneurship & Society at HIIG. In her research she investigates risks and potentials of AI in people management. Currently, she is staying at the policy lab Denkfabrik Digitale Arbeitsgesellschaft as a research fellow.

Miriam Klöpper is currently a researcher at FZI Forschungszentrum Informatik. She is interested in the social impact of digitalisation, especially in the context of workplace technologies such as (predictive) people analytics.

the use of people analytics can impair the experience of working remotely by invading privacy and impeding worker autonomy.

PEOPLE ANALYTICS AND WORKS COUNCILS

Against the backdrop of these developments, both management and works councils are facing new challenges. Whether people analytics tools serve the interests of employees depends on how they are designed, implemented and put into practice – ultimately, the involved actors need to define meaningful metrics so that they can draw informed conclusions from the data. If there is a bias in the training data, it will most likely be perpetuated by the algorithm (Raghavan, Barocas, Kleinberg & Levy, 2020). To protect employees' privacy, it is vital to have an anticipatory process when choosing and implementing a people analytics tool. Works councils should thus be involved in the process at all times, including at the early stages of choosing a vendor. Nevertheless, even the most dedicated works councils may struggle to comprehensively evaluate people analytics and the variety of implications these tools (may) have for employees. The general need to understand and use digital technologies has grown throughout the past year. With it, the range of training opportunities for works councils, often organised by unions, has expanded. Taking courses to understand the implications of algorithm-driven technologies such as people analytics is a necessary first step. However, because these tools can often be described as *black box* systems that are intricate and include mechanisms that are not transparent to the user, more rigorous training and educational resources are needed (AlgorithmWatch, 2020).

WHAT'S AHEAD?

People analytics tools offer a wide range of helpful features. Yet, without guidelines and constant monitoring of the outcome of algorithmic decision-making, these tools may cause severe problems such as discrimination. It is essential to ensure the transparency of the systems and of any sort of data collection, so that employees can make informed decisions when using them. Works councils play a key role in the deployment of people analytics, as, by German law, these systems cannot be implemented and used without the prior agreement of the individual employee or a company agreement (AlgorithmWatch, 2020). There is, at present, insufficient research about the social and ethical implications of people analytics, especially in the context of remote work. The limited knowledge on the topic makes it difficult for works councils and employees to educate themselves about the systems.

Further research in this area that provides low-threshold information is therefore urgently needed. ◆

REFERENCES

AlgorithmWatch (2020). *Positionen zum Einsatz von KI im Personalmanagement. Rechte und Autonomie von Beschäftigten stärken – Warum Gesetzgeber, Unternehmen und Betriebsräte handeln müssen*. https://algorithmwatch.org/de/auto-hr/positionspapier

Avolio, B. J., Sosik, J. J., Kahai, S. S., & Baker, B. (2014). E-leadership: Re-examining transformations in leadership source and transmission. *The Leadership Quarterly, 25*(1), 105–131. https://doi.org/10.1016/j.leaqua.2013.11.003

Banholzer, N., Feuerriegel, S., Fleisch, E., Bauer, G. F., & Kowatsch, T. (2021). Computer Mouse Movements as an Indicator of Work Stress: Longitudinal Observational Field Study. *Journal of Medical Internet Research, 23*(4). https://doi.org/10/gjxrvf

BMFSFJ – Bundesministerium für Familie, Senioren, Frauen und Jugend (2017). *Digitale Vereinbarkeit: Home-Office und mobiles Arbeiten: Home-Office und mobiles Arbeiten - eine Toolbox für Unternehmen und Beschäftigte mit Familienaufgaben*. https://www.bmfsfj.de/bmfsfj/service/publikationen/digitale-vereinbarkeit-home-office-und-mobiles-arbeiten-118754

Cristea, I. C., & Leonardi, P. M. (2019). Get Noticed and Die Trying: Signals, Sacrifice, and the Production of Face Time in Distributed Work. *Organization Science, 30*(3), 552–572. 10.1287/orsc.2018.1265

Estévez-Mujica, C. P., & Quintane, E. (2018). E-mail Communication Patterns and Job Burnout. *PLOS ONE, 13*(3), 1–25. 10.1371/journal.pone.0193966

Fleming, P., & Spicer, A. (2004). 'You Can Checkout Anytime, but You Can Never Leave': Spatial Boundaries in a High Commitment Organization. *Human Relations, 57*(1), 75–94. https://doi.org/10.1177/0018726704042715

Grunau, P., Ruf, K., Steffes, S., & Wolter, S. (2019). *Mobile Arbeitsformen aus Sicht von Betrieben und Beschäftigten: Homeoffice bietet Vorteile, hat aber auch Tücken* (IAB-Kurzbericht, 11/2019). https://www.iab.de/194/section.aspx/Publikation/k190604j02

Jarrahi, M. H., Newlands, G., Lee, M. K., Wolf, C. T., Kinder, E., & Sutherland, W. (2021). Algorithmic management in a work context. *Big Data & Society, 8*(2). https://doi.org/10.1177/20539517211020332

Leonardi, P. M., Treem, J. W., & Jackson, M. H. (2010). The Connectivity Paradox: Using Technology to Both Decrease and Increase Perceptions of Distance in Distributed Work Arrangements. *Journal of Applied Communication Research, 38*(1), 85–105, https://doi.org/10.1080/00909880903483599.

Martinez, D. E. (2011). Beyond disciplinary enclosures: Management control in the society of control. *Critical Perspectives on Accounting, 22*(2), 200–211. https://doi.org/10.1016/j.cpa.2010.06.016

Möhring, K., Naumann, E., Reifenscheid, M., Weiland, A., Blom, A. G., Wenz, A., Rettig, T., Lehrer, R., Krieger, U., Juhl, S., Friedel, S., Fikel, M., & Cornesse, C. (2020). *Die Mannheimer Corona-Studie: Schwerpunktbericht zur Nutzung und Akzeptanz von Homeoffice in Deutschland während des Corona-Lockdowns*. https://madoc.bib.uni-mannheim.de/55628

Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 469–481. https://doi.org/10.1145/3351095.3372828

Sewell, G., & Taskin, L. (2015). Out of Sight, Out of Mind in a New World of Work? Autonomy, Control, and Spatiotemporal Scaling in Telework. *Organization Studies, 36*(11), 1507–1529. 10.1177/0170840615593587

Staab, P., & Geschke, S.-C. (2020). *Ratings als arbeitspolitisches Konfliktfeld: Das Beispiel Zalando*. https://www.boeckler.de/de/faust-detail.htm?sync_id=HBS-007864

Taskin, L., & Devos, V. (2005). Paradoxes from the Individualization of Human Resource Management: The Case of Telework. *Journal of Business Ethics, 62*(1), 13–24. https://doi.org/10.1007/s10551-005-8710-0

Tursunbayeva, A., Di Lauro, S., & Pagliari, C. (2018). People analytics — A Scoping Review of Conceptual Boundaries and Value Propositions. *International Journal of Information Management*, 43, 224–247. 10.1016/j.ijinfomgt.2018.08.002


60
10
50
20
40
30
5
10
55
50
45
40
35
30
25
MADE IN USSR

"[T]here is a fundamental tension between competition and cooperation when internet networks interconnect."

THE *DIGITALER SALON* AS SEEN THROUGH THE EYES OF A JOURNALIST

AN INTERVIEW WITH KATJA WEBER BY NATASHA VUKAJLOVIC

Lost in Regulation, Machine Gun Learning, App's Anatomy, Click Chat Love, Swipe for President, The Troll next Door... under these titles, we have been publicly debating the impact of digitalisation on society on a monthly basis at Alexander von Humboldt Institute for Internet and Society (HIIG) and other venues. In our talk series *Digitaler Salon*, we have shed light on digital phenomena, discussed the advantages and disadvantages of technical changes and provided food for thought – for almost a decade now. Since her first episode on 30 April 2014, our front woman Katja Weber has moderated more than 70 panel discussions for this format. Who is a better person to ask about striking trends, remarkable guests and moments to remember at *Digitaler Salon*? The conversation was led by Natasha Vukajlovic who has cordinated the event series at HIIG in the past two years.

Natasha: Let's take a little leap back in time: the first panel discussion you moderated at HIIG was in 2014, and then you moved with us from the live discussion and "Hörsaal" radio show formats at Bebelplatz to live streaming at the Digitaler Salon in Französische Straße and, during the pandemic, to video conferencing. How did you feel about these changes? How did you adapt to the changes?

Katja: The most significant change for me was when Digitaler Salon became truly digital due to the pandemic. Before that, the salon was clearly always an on-site event. I thought the room at Bebelplatz was great, with the small kitchenette and the students walking past. It was a very low-threshold approach, where you could stumble in by mistake and then find yourself in a flat-sharing atmosphere. Even when the discussion was subsequently broadcast on Deutschlandfunk Nova, that was not decisive for how I conducted the discussion on site. It was first and foremost an event for the audience in the room. That's still what the salon is at the events in Französische Straße, but hardly anyone will drop in by chance. Those who go to the salon now are looking for it specifically. In this respect, it's a more set framework,

which perhaps fits in with the fact that HIIG and the salon have now established themselves over a period of time.

The biggest change was, of course, the pandemic – and the fact that we went digital in every respect, with guests joining in, streaming and being online only for several months. And lo and behold, it worked enormously well! Later, I also found hybrid formats – some guests connected, others on site – to be very successful. You did a great job, turned things around quickly and made this exchange possible. Many other organisers cancelled everything at first. You have an audience that is eager to discuss things, and they went along with you and, if they can't do it in person, they also get involved online by asking questions. It's good that we were able to offer them content and not just Netflix during the period of confinement.

We have talked primarily about the impact of digitalisation in various areas of life and society. How has digitalisation changed the Digitaler Salon and your work over the years?

As I just described, this is primarily due to the primacy of online-only formats from March 2020. Of course, we now also use another tool, Slido, to collect questions from the audience. But my preparation or the execution itself, the conversation with guests and audience has not changed for me. With regard to the audience: I'm pleased that there are regular guests. They are so interested, they come whenever they can, regardless of the topic. But then there are also people who came very regularly for a long time and then suddenly stopped coming. Maybe they moved away?

If we're talking about trends that we've seen play a role on the panel, then Google Glass as a product and the sharing economy as a concept come directly to mind. In the mid-10s, data glasses were the privacy bugbear – now they seem to have pretty much faded into obscurity or niches. The sharing economy, on the other hand, is always evolving. I remember how we had an Uber Germany manager as a guest in 2015 – he was really grilled by the audience. In the meantime, we're talking about data donations, cooperatives – there's still a lot of interest in this topic. Oh, another trend: dating platforms. In 2016, the salon on this topic was so overrun that our event hall was not big enough and you had to broadcast via stream in a second room. The audience had an enormous number of questions about Tinder and similar services – of course the questioners always added "I'm asking for a friend". When we talked about it this year in May, this excitement was gone, online dating is no longer shameful but just a part of everyday life.

What motivates you to moderate the Digitaler Salon over such a long period of time?

When it started with you and me in 2014, I didn't realise how long the liaison would last. But digitalisation is so comprehensive that no area remains untouched. We're not just talking about AI, open source or the work of the future, but also about love, mourning, literature – everything is changing. Working that out and recording it is a lot of fun. And even if a topic comes up again after a few years, it is not the same but has significantly changed, as I just described in relation to dating. A lot of factors come together: good topics. You invite great guests who then also have something to say seriously. We have a great audience that contributes a lot and drills down into the topics that are important to them. You at HIIG have a knack for finding dedicated people who do everything they can in the preparation, organisation and running of the event to make it a success, whether they are permanent employees or student assistants. The technical team on the evening of the event is cool – they always bring us to Alex TV with great composure. And I love the titles you come up with. "Machine Gun Learning". "The Zooming Classroom". Or, my all time favourite: "Dem Hölderlin sein Hashtag". Great art.

Have you developed a routine for preparing for the discussions over the years? And if so, what does it look like?

I read up broadly on the press coverage of our topics and guests beforehand. You send me a briefing with what is important to you. And on this basis, I then figure out what questions I expect to discuss. Sometimes this works out very well and my guideline works out one hundred percent. And sometimes, what happens – especially at the events in person – is that the audience wants to find out about just one aspect but absolutely wants more depth, which is great. Then I put my preparation aside and continue along these interests.

What was your favourite episode? And who was the most remarkable guest and why?

It would be unfair to single out one premium guest here. Often, it's only together that the whole thing comes together. In the salon, the person sitting on the right doesn't have to rhetorically beat up the person sitting on the left, and in the end, one of them leaves as the champ. Ideally, they look at a topic from different perspectives and then complement each other. Like the Zooming Classroom Digitaler Salon in January

2021, when the schools were closed: a Berlin student, an education researcher and a data protection activist dad – they have something to say to each other, without mud-wrestling. That gets me going – and I'm guessing that the same applies to the viewers, too.

My favourite episode is also hard for me to answer. What I always like is when a topic that seems somehow brittle – art and digitalisation or faith and digitalisation - suddenly opens up during the event, becomes totally insightful and poses lots of exciting questions for me and the audience. Even though I might have thought beforehand: Let's see if this works.

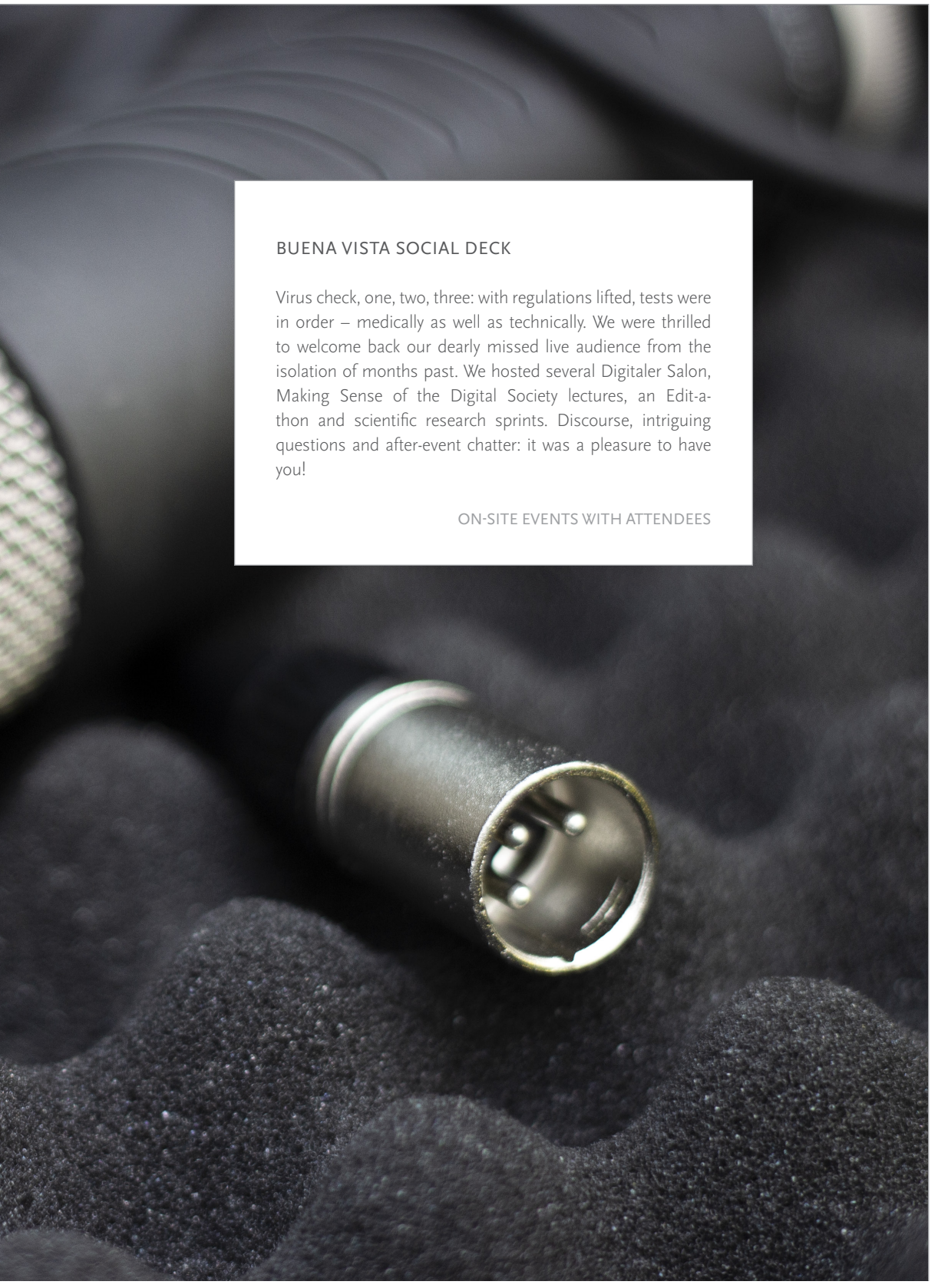
How do you describe the Digitaler Salon to people who don't know the format?

I still think the subtitle "Questions about the networked present" is precise and apt. It gets to the heart of the matter. We discuss all topics – from the point of view of digitalisation.

A look into the future: How do you envision the Digitaler Salon in another 10 years?

Ideally, it should be just as diverse as it is now. It would be nice if we had fewer occasions to talk about hate speech, populism and propaganda online. And if one of our perennial hot topics – data protection – had a more positive image. But I think I'm already noticing the right vibes. ◆


DIGITALER SALON



BUENA VISTA SOCIAL DECK

Virus check, one, two, three: with regulations lifted, tests were in order – medically as well as technically. We were thrilled to welcome back our dearly missed live audience from the isolation of months past. We hosted several Digitaler Salon, Making Sense of the Digital Society lectures, an Edit-a-thon and scientific research sprints. Discourse, intriguing questions and after-event chatter: it was a pleasure to have you!

ON-SITE EVENTS WITH ATTENDEES

D

JAYA KLARA BREKKE AND ARON FISCHER

Digital scarcity

Digital scarcity is a credibly maintained limitation – imposed by software – of digital information, goods or services that may be accessed and used entirely digitally.

Some of the earliest uses of the term digital scarcity stem from the early 2000s and describe the scarcity of access to IT resources and the underlying physical resources that computers and networks rely on—i.e. "the scarcity of the digital"(Weinberger, 2003; Hammersley, 2003; Chaudhry & Shipp, 2005). As internet access has become more widespread, and as an increasing amount of content is consumed digitally, the usage of the phrase has shifted. Digital information is by nature not scarce or rivalrous; it can be copied and shared at next to no cost, with no reduction in availability or quality. This in turn imposes severe limitations on online copyright enforcement. In this context, digital scarcity refers to the imposition of limits and conditions on the availability of and access to digital content, aimed at protecting business models that depend on scarcity.

The rise of the internet led to movements of digital activists seeking to open up access to information entirely ("information wants to be free"). These movements often clashed with intellectual property and copyright-based industries (Dahlstrom et al., 2006; Swartz, 2016), and these clashes in turn informed much of the development of peer-to-peer systems that would enable the circumvention of the copyright industry and allow free access to information (Oram, 2001; Andersson, 2011). It was argued that nothing digital is genuinely scarce and that any imposed scarcity is not just artificial but also objectionable. If we look beyond mere data, there are digital resources that are inherently limited, such as bandwidth or short domain names. Sometimes the creation of digital scarcity is accidental and its maintenance is due to a failure in governance. A prominent example is the dearth of IPv4 (Internet Protocol version 4) addresses (Rodriguez, 2012).

DIGITAL SCARCITY IN THE AGE OF BLOCKCHAINS

As the copyright battles of the 1990s and early 2000s made clear, maintaining digital data scarcity by preventing copying is nearly impossible. However, establishing referential scarcity, where references are ledger/database entries (and the referents are anything from cryptocurrencies to cryptokitties), is possible as long as it can be credibly established that the scarcity will be maintained and the rules adhered to. The crucial aspect of referential scarcity is not control over data availability but control over manipulation of the data in question. With the invention of Bitcoin, referential scarcity could be established without the need for a central entity to enforce it. Cryptocurrencies are not the first databases with finite number entries, but they are the first in which changes to the entries cannot be forced by the entities providing the computing infrastructure. The notion

that centralised control over a database is necessary to ensure digital scarcity has thus been overturned.

As more advanced and general-purpose blockchain networks such as Ethereum have appeared, the scope for scarce ledger entries has grown. Aside from the scarcity of cryptocurrencies and currency-like *tokens*, a new class of *unique digital items*, known as non fungible tokens or NFTs, have appeared. These range from formal claims of ownership over a real-world (offline) asset, to purely digital collectibles (see, for example, Serada, Sihvonen, and Harviainen, 2020). At the time of writing, the culture around blockchains is still young, and it remains highly politicised and polarised. This polarisation has contributed to the confusion surrounding digital scarcity specifically as it relates to ideas of value. Proponents of Bitcoin in particular argue that it is the limited supply of bitcoins (and that alone) that gives them intrinsic value whereas supporters of other blockchains (such as Ethereum, Cardano, Polkadot) argue that the utility of the network, its extrinsic value, is far more important.

Digital scarcity describes a credibly maintained limitation – imposed through software – of digital information, goods or services that may be accessed and used entirely digitally. This includes limitations to entries in a ledger or database (including cryptocurrency entries in a blockchain or top-level domains in the Domain Name System), as well as limitations in access to computing resources such as network addresses, bandwidth or (again in the context of blockchains) transactions-per-second, wherever these limits go beyond the physical limits imposed by hardware. The motivations for engineering digital scarcity tend to be in order to support business models that profit from scarcity or uniqueness in the digital realm. ◆

The *Glossary of decentralised technosocial systems*, edited by Valeria Ferrari for Internet Policy Review, is an interdisciplinary glossary on peer-to-peer, user-centric and privacy-enhancing decentralised technologies. In order to tackle the existing gap in shared semantics, this glossary converges the efforts of experts from various disciplines to build a shared vocabulary on the social, technical, economic, political aspects of decentralised, distributed or sovereign technologies.

policyreview.info/glossary

REFERENCES

Andersson, J. (2011). The origins and impacts of the Swedish file-sharing movement: A case study. *Critical Studies in Peer Production (CSPP), 1*(1), 1–18. http://urn.kb.se/resolve?urn=urn:nbn:se:sh:diva-22494

Chaudhry, V., & Shipp, T. (2005). Rethinking the digital divide in relation to visual disability in India and the United States: Towards a paradigm of 'Information Inequity. *Disability Studies Quarterly, 25*(2). https://doi.org/10.18061/dsq.v25i2.553

Hammersley, B. (2003, July 10). Radio active revolution. *The Guardian*. https://www.theguardian.com/technology/2003/jul/10/onlinesupplement2

Oram, A. (Ed.). (2001). *Peer-to-peer: Harnessing the benefits of a disruptive technology* (1st ed.). O'Reilly.

Rodriguez, R. (2012, November). IPv4 scarcity [Blog post]. *Internet Society*. https://www.internetsociety.org/blog/2012/11/ipv4-scarcity

Roemer, R., Schear, N., Dahlstrom, D., & Farrington, N. (2006). Piracy in the digital age. *University of Washington, Computer Science & Engineering*. https://courses.cs.washington.edu/courses/csep590a/06au/projects/digital-piracy.pdf

Serada, A., Sihvonen, T., & Harviainen, J. T. (2020). CryptoKitties and the new ludic economy: How blockchain introduces value, ownership, and scarcity in digital gaming. *Games and Culture*. https://doi.org/10.1177/1555412019898305

Swartz, A., & Lessig, L. (2016). *The boy who could change the world: The writings of Aaron Swartz*. Verso.

Weinberger, D. (2003, March 13). The myth of interference. *Salon*. https://www.salon.com/2003/03/12/spectrum

MAX VON GRAFENSTEIN AND ELIAS BELGACEM

Effective regulation through design

How can the ePrivacy Regulation draft be aligned with the actual needs of users? This article examines the ambiguous interplay of the new regulation with the GDPR tracking technologies in personalised internet content and the data-protection-by-design approach.

EU legislators are currently negotiating the ePrivacy Regulation in the trilogue procedure. Given the latest draft released by the council, we believe it is flawed. This flaw lies in the ambiguous relationship between the ePrivacy Regulation and the GDPR, which calls into question the applicability of several decisive provisions of the GDPR (e.g. the data-protection-by-design approach and co-regulation instruments).

The communications sector is characterised by two key aspects: the rapid pace of technological development and the dependency of users on the trustworthiness of communication providers. Since third parties mediate data subjects' communication, these data subjects can exercise limited control over their privacy, freedom, equality, etc. Based on our interdisciplinary research, which focuses on personalised content and tracking technologies, we observe that the current draft does not provide a level of protection that could be considered effective in meeting the needs of communications users.

For instance, the latest draft of the ePrivacy Regulation requires the consent of data subjects as an important regulatory mechanism (see Art. 4a). However, the consent requirement runs the risk of being ineffectual because consent alone cannot solve the problem of third-party dependence. Whether the providers of communication media adhere to the conditions of consent depends on the trustworthiness of the providers. The resulting consent fatigue (Choi et al., 2018) is also problematic. It is a weariness that results from the frequency with which data subjects' consent is requested and the way in which the actual content design is presented. Providers must therefore implement consent so that data subjects can effectively make an informed choice and not just give consent out of frustration or fatalism. This regulatory goal applies in particular to cookies and other tracking technologies. Art. 25 sect. 1 GDPR is the key data protection provision and focuses explicitly on the effective implementation of protection measures like informed consent. The data-protection-by-design approach requires (data) controllers to effectively implement legal provisions in the technical and organisational design of the data processing. Moreover, the certification mechanisms and codes of conduct under Art. 40 to 43 GDPR ensure that controllers act on the conditions of consent given by data subjects, since data subjects have little capacity to verify this themselves. Thus, the ePrivacy Regulation must clarify that these GDPR provisions are applicable to the processing of personal data concerning communication.

To test our hypothesis regarding whether the current draft per se sufficiently protects communication

media users' privacy expectations with respect to tracking technologies, we first conducted a qualitative user study. Based on these results, we sketched out some initial design drafts for transparency and control options that may better meet data subjects' demands. Finally, we analysed the current draft of the ePrivacy Regulation to determine to what extent it supports designs such as ours or if it would be necessary to fall back on the aforementioned provisions of the GDPR.

USERS' NEEDS REGARDING PERSONALISED CONTENT AND TRACKING TECHNOLOGIES

We conducted 20 qualitative interviews. The examples discussed with our interviewees ranged from advertising to news, prices and electoral advertising – all personalised. The tracking technologies ranged from cookies to logins to newer techniques (e.g. fingerprinting). Among the themes we discussed with our interviewees, the following are worth more detailed consideration:

The value of personalised content: users did not disapprove of personalised content in general but rather recognised it as an important feature for businesses and users themselves, since it allows individuals to discover new products or find better prices etc.

"Consent fatigue" and "creepy moments": however, users often experienced "creepy moments" (Omer & Poloentsky, 2013) and "consent fatigue" (Choi et al., 2018) due to a lack of public transparency on the process behind content personalisation, users' own ignorance of the process, and the need to navigate dark pattern manipulation. Creepy moments often arose from unexpected displays of content that a user attributed to their profile but could not explain.

Opaqueness of profiles: most users did not know – despite wanting to know – why they see the content they see (i.e. how the content gets personalised). Given the example of personalised ads, users did not understand why they were seeing the ads they were seeing – that is, based on which attributed interests and on which of their collected personal data.

Users' ignorance of tracking technologies: another reason for creepy moments was a lack of understanding of how they are identified online and for whom the content is being personalised (e.g. for the user, their family, their flatmates). Users who shared one technical identity (e.g. via a cookie) often wondered why

they could track each other's surfing behaviour via the personalised content displayed to them.

Control and deceptive design: once users understood the functionality of tracking technologies, they preferred opt-in to opt-out solutions. They also wanted a general toggle button from personalised content to non-personalised content in order to understand what is visible to the general public. However, users were well aware of the use of manipulative cookie banner design and considered it very annoying.

Uncertainty about concrete solutions: while some remained fatalistic or uninterested, the majority of our interviewees still wanted better transparency and control. More importantly, when asked how transparency and control could be improved specifically, most participants were quickly overwhelmed. Thus, the proposals for solutions remained largely superficial, although those solutions varied, ranging from better-tailored advertising to better protection of privacy and data trading.

To give users a more concrete idea of how transparency and control options could be better designed, we drew up our own initial designs. These designs also allowed us to examine whether the current draft of the ePrivacy Regulation supports such designs.

DESIGNING EFFECTIVE TRANSPARENCY AND CONTROL ARCHITECTURES FOR COOKIE BANNERS

In our design study, we decided to use privacy icons as eye-catchers and a layered approach to present privacy notices and control options. The use of privacy icons and a layered approach is not mentioned in the current draft of the ePrivacy Regulation. However, the effectiveness of our approach could be tested and enforced in practice based on Art. 25 sect. 1, Art. 5 sect. 1 lit. a and Art. 12 GDPR.

Further, we decided that our privacy icons should form a data protection seal within the meaning of Art. 42 GDPR. The idea behind this is that the use of icons must be secured by appropriate certification procedures in order to avoid possible misuse (= false declaration) and thus ensure the trustworthiness of the seal. The current draft of the ePrivacy Regulation does not mention certification mechanisms, however, these mechanisms are necessary to ensure that controllers adhere to the processing conditions described in their privacy statement.

To better meet the needs expressed by our interviewees, our data protection strategy also aims at minimising individuals' risk of being manipulated by personalised content. It does so by offering them usable information and control over the profiles on which the personalised content is based. By clicking on such a processing purpose on level 2, like for personalised advertising, users can go directly to their profile in level 3. Level 3 informs users about their profile: for instance, users can see, adjust and delete the interests that are attributed to them. They can see, complete and delete the personal data that has been collected in order to assign the resulting attributed interests in the personalised advertisement at hand. They can thus understand the content personalisation they experience based on their concrete usage context and maintain their autonomy in their usage decisions. The information is also intended to avoid the so-called creepy moments mentioned in our qualitative research. However, while the current draft of the ePrivacy Regulation does not require the information given to the data subjects to have a certain level of detail nor for the information to be placed in a certain usage context, such designs could be developed on the basis of Art. 25 sect. 1 GDPR. If such designs are empirically proven to protect users more effectively, this would constitute the new state of the art which all controllers have to consider.

MORE EFFECTIVE REGULATION THROUGH HUMAN-CENTRED DESIGN

At least with respect to personalised content and tracking technologies, the current draft hardly meets users' needs. It regulates, for instance, whether consent is necessary. However, it does not address any of the subsequent questions of how users should be informed and how accessible the means of consent should be to enable users to make informed decisions related to tracking and associated purposes. Worse, we consider the effective implementation of transparency measures and control options unlikely if the current draft does not clarify its exact interplay with the GDPR.

In any case, applying the data-protection-by-design approach to the ePrivacy Regulation means that pointless cookie banners may soon be history. If so, legislators would not have to dictate what information or control architectures should look like from a user's perspective. In addition, the rapid pace of development in the communications sector would otherwise prove extremely challenging for legislators unless a data-protection-by-design approach is applied to the ePrivacy Regulation. Legislators can certainly make individual specifications, as in Article 4a of the draft. However, they should be careful not to make too

continue reading on page 156 ▸▸

THIS IS AN ARTICLE BY **MAX VON GRAFENSTEIN AND ELIAS BELGACEM**

This article is an abridged version of a position paper by Max von Grafenstein, Julie Heumüller, Timo Jakobi, Elias Belgacem and Patrick Smieskol published on *Zenodo*.

Max von Grafenstein is professor for digital self-determination at Berlin University of the Arts (UdK) in association with the Einstein Center Digital Future (ECDF). He is particularly interested in the "data-protection-by-design" approach, which aims to incorporate the requirements of data protection law into the design of data-based products and companies.

Elias Belgacem is research assistant in Max von Grafenstein's team. In this context, he analyses how to make movement data available for sustainable urban mobility from a data protection perspective. He is also the founder and chair of the Euro-Mediterranean Legal Center (EMLC).

many specifications. These can quickly become obsolete or ineffectual in light of rapid advances within the communications sector. Since data protection by design requires controllers to implement protection effectively by taking the state of the art into account, the approach offers a two-fold advantage: legislators provide adequate protections for data subjects while leaving the door open to build upon existing protection measures as the communication sector evolves.

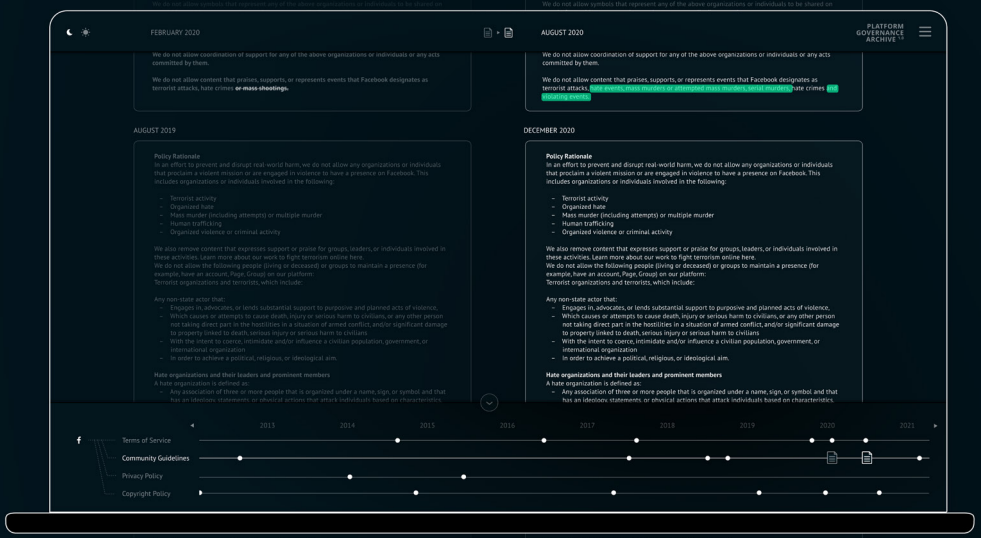
This leads us to the most fundamental aspect of our criticism: how can legislators design laws that more effectively address user needs? How can they avoid ambiguities in laws that jeopardise the effective implementation of protection measures in practice? In our opinion, legislators can achieve this goal by expanding their legislative methods. While legislation should still draw from legal considerations involved in the legislative information process, we suggest that this process would benefit considerably from empirical studies and design methods such as those presented in this paper. Accordingly, legislators could test the effects of their regulation in practice, thereby increasing the effectiveness and the rationality of laws (Hoffmann-Riem & Fritzsche, 2009). In conclusion, we argue for more evidence-based lawmaking through human-centred design. ◆

REFERENCES

Choi, H., Park, J., & Jung, Y. (2018). The Role of Privacy Fatigue in Online Privacy Behavior. *Computers in Human Behavior*, 81, 42. 10.1016/j.chb.2017.12.001

Hoffmann-Riem, W., & Fritzsche, S. (2009). Innovationsverantwortung: Zur Einleitung. In W. Hoffmann-Riem & M. Eifert (Ed.), *Innovationsverantwortung* (Innovation und Recht, III, 11-41). Duncker & Humblot.

Omer, T., & Poloentsky, J. (2013). A Theory of Creepy: Technology, Privacy and Shifting Social Norms. *Yale Journal of Law and Technology, 16*(1), 59–102. https://digitalcommons.law.yale.edu/yjolt/vol16/iss1/2


FEBRUARY 2020
We do not allow coordination of support for any of the above organizations or individuals or any acts committed by them.
We do not allow content that praises, supports, or represents events that Facebook designates as terrorist attacks, hate crimes or mass shootings.
AUGUST 2020
We do not allow coordination of support for any of the above organizations or individuals or any acts committed by them.
We do not allow content that praises, supports, or represents events that Facebook designates as terrorist attacks, hate events, mass murders or attempted mass murders, serial murders, hate crimes and violating events.
PLATFORM GOVERNANCE ARCHIVE v1.0
AUGUST 2019
Policy Rationale
In an effort to prevent and disrupt real-world harm, we do not allow any organizations or individuals that proclaim a violent mission or are engaged in violence to have a presence on Facebook. This includes organizations or individuals involved in the following:
– Terrorist activity
– Organized hate
– Mass murder (including attempts) or multiple murder
– Human trafficking
– Organized violence or criminal activity
We also remove content that expresses support or praise for groups, leaders, or individuals involved in these activities. Learn more about our work to fight terrorism online here.
We do not allow the following people (living or deceased) or groups to maintain a presence (for example, have an account, Page, Group) on our platform:
Terrorist organizations and terrorists, which include:
Any non-state actor that:
– Engages in, advocates, or lends substantial support to purposive and planned acts of violence.
– Which causes or attempts to cause death, injury or serious harm to civilians, or any other person not taking direct part in the hostilities in a situation of armed conflict, and/or significant damage to property linked to death, serious injury or serious harm to civilians
– With the intent to coerce, intimidate and/or influence a civilian population, government, or international organization
– In order to achieve a political, religious, or ideological aim.
Hate organizations and their leaders and prominent members
A hate organization is defined as:
– Any association of three or more people that is organized under a name, sign, or symbol and that has an ideology, statements, or physical actions that attack individuals based on characteristics,
DECEMBER 2020
Policy Rationale
In an effort to prevent and disrupt real-world harm, we do not allow any organizations or individuals that proclaim a violent mission or are engaged in violence to have a presence on Facebook. This includes organizations or individuals involved in the following:
– Terrorist activity
– Organized hate
– Mass murder (including attempts) or multiple murder
– Human trafficking
– Organized violence or criminal activity
We also remove content that expresses support or praise for groups, leaders, or individuals involved in these activities. Learn more about our work to fight terrorism online here.
We do not allow the following people (living or deceased) or groups to maintain a presence (for example, have an account, Page, Group) on our platform:
Terrorist organizations and terrorists, which include:
Any non-state actor that:
– Engages in, advocates, or lends substantial support to purposive and planned acts of violence.
– Which causes or attempts to cause death, injury or serious harm to civilians, or any other person not taking direct part in the hostilities in a situation of armed conflict, and/or significant damage to property linked to death, serious injury or serious harm to civilians
– With the intent to coerce, intimidate and/or influence a civilian population, government, or international organization
– In order to achieve a political, religious, or ideological aim.
Hate organizations and their leaders and prominent members
A hate organization is defined as:
– Any association of three or more people that is organized under a name, sign, or symbol and that has an ideology, statements, or physical actions that attack individuals based on characteristics.
2013
2014
2015
2016
2017
2018
2019
2020
2021
Terms of Service
Community Guidelines
Privacy Policy
Copyright Policy

THE RISE OF PLATFOMS AND THEIR GOVERNANCE

AN INTERACTIVE GENEALOGY OF PLATFORM POLICIES

How have Facebook's policies regarding misinformation changed over time? When did Twitter include the line "Twitter's purpose is to serve the public conversation" in their Community Guidelines? And when did Facebook remove a provision that prohibits the posting of "false information" from their Terms of Service? What cultural and political norms are embedded in the ever-more extensive rulebooks that internet platforms have formulated? And what general patterns and trends run through the historical evolution of these policies?

Over the last 15 years, internet platforms have emerged as key social and technical institutions for communication and interactions, and the way in which they govern user activities has tremendous consequences for how our increasingly digital society is organised. Growing controversies around issues such as hate speech and misinformation point to the role and responsibility of platforms for the regulation of public speech and communication dynamics. The platforms have reacted by gradually abandoning their self-positioning as neutral technical intermediaries and by formulating increasingly extensive sets of rules that govern which kind of content and conduct is allowed and prohibited on their services.

It is these rules that are the subject of the Platform Governance Archive, an open access repository of platform policies. The interactive website addresses the growing need for a systematic study of platform policies and their historical evolution. It allows citizens, researchers and journalists alike to understand how this governance by platforms has changed over time. The archive is currently based on a dataset of three key policy types by four major platforms that were collected and curated via a combination of automated and manual approaches. It will be extended in the future to establish a common data repository and shared data infrastructure for the growing field of platform governance research. ◆

Explore the archive and give feedback for its future development.

pga.hiig.de

THOMAS SCHILDHAUER, ANNIKA
ULICH AND PIA WINCKLER

Tech & aging: how to enable independent living with digital innovations

The health and care sector faces a myriad of irresolvable challenges – or so it seems. The questions are manifold: how can your older relatives age on their own terms, in their own home? How can understaffed healthcare facilities support their employees? How can changing demographics and changing needs be addressed? The digitalisation of care can provide important answers.

Take a look at Frida Müller's everyday life: Frida is 86 years old and lives in an assisted living facility, where she has her own apartment. She receives help getting dressed in the mornings and her granddaughter visits her on weekends. As she has got older, she has become more afraid of falling when she goes for a walk or runs errands – in situations like these, she is concerned that she will not be able to get help quickly enough. The nurses are often short on time, so Frida often keeps these worries to herself. She also prefers not to burden her friends and family with medical questions. To adequately respond to challenges like these on a systemic level, we need to bring together digital technologies and the perspectives of the elderly, their relatives, caregivers and the care sector along with its interdisciplinary researchers and diverse stakeholders. Key questions include: are there digital solutions that could support Frida when she feels anxious and left alone with her worries? What would these solutions look like? How would she use them?

The creation of a comprehensive, interdisciplinary research environment is a necessary condition for designing and researching digital care and health technologies that adapt to people and their social environments and focus on the supporting and hindering factors for the usage of technology in caring for the aging population. This article highlights the contemporary problems of aging and health as well as digitalisation in these areas. There are two key ways of approaching such issues. The first concerns the user-centred design of digital solutions to relieve the burden on professional and family carers. The second is the *Learning Lab* research concept, which has a multi-disciplinary approach and constitutes the neighbourhood as a space where user-centred design is implemented, observed and evaluated. Against this backdrop, the Digital Urban Center for Aging and Health (DUCAH) initiative was launched – based on the principles of user-centred design and the Learning Lab approach.

CURRENT AGING DEBATES

The current situation in the care sector is characterised by the effects of demographic change, the shortage of skilled workers, a high level of regulation and the emergence of new digital technologies. Elderly people and their families have to make difficult decisions to safeguard their well-being and happiness now and in the future. This includes the desire to maintain as much independence as possible and to stay in their own home. It can be hard for ever-smaller families to make these wishes a reality and take on the

responsibilities of caring for their relatives. Nursing homes or nursing services come with their own challenges. Frida Müller might have been very unwilling to relocate to her assisted living facility, uncomfortable with depending on strangers and the financial strain might have been difficult to manage. Especially in the nursing sector, employees take on large workloads with small wages because facilities tend to be understaffed, which can impact the quality of care they are able to deliver. While Frida Müller receives help, she does not feel adequately supported. Neither the nursing staff nor her family have the capacity to be there for her in her everyday life when she has a question or needs encouragement. At the same time, Frida would like to live as independently as possible.

In addition, the care sector is becoming more and more digital as new technologies emerge. Unfortunately, these digital solutions are not proven to be ready to use – often they meet neither the specific expectations of the elderly nor the wider living environment in which they find themselves. The elderly must adapt to complex and inflexible technical systems, which are often not unified, consistent or seamless. This causes various problems and challenges: first, people may be unable to adapt their living conditions to utilise these digital solutions and consequently have to move to a nursing home. They may misuse the technology or seek to work around it to avoid problems which would require asking for help. Second, most research settings in this field are similarly dysfunctional – they are highly artificial and built around the technologies to be researched rather than focusing on the desired outcome and the people. They thus resemble technology showrooms rather than the living environment in which they eventually will be used.

New forms of cooperation need to be created and more relevant stakeholders need to be integrated into development processes to meet these challenges. In the fields of aged care, healthcare, digital technology, district planning, neighbourhood development and housing construction, these problems can be tackled together and intersectoral solutions can be developed. In Frida Müller's case, for example a concierge service could give her support without limiting her independence. Katja Gast from Diakonie Deutschland and Jürgen Albert from German Medical Association work on this solution at DUCAH: "We want our concierge service to act as an anchor for people in their neighbourhoods on the topic of health and to empower them to become (digitally) enabled and involved."

continue reading on page 164 ▸▸

THIS IS AN ARTICLE BY **THOMAS SCHILDHAUER, ANNIKA ULICH AND PIA WINCKLER**

This article was first published on 27 January 2022 on the *Digital Society Blog* of the Alexander von Humboldt Institute for Internet and Society (HIIG). Find out more about aging, independent living and technology on *ducah.de*.

Thomas Schildhauer is initiator of the Digital Urban Center for Aging & Health (DUCAH) and one of HIIG's research directors. He holds a professorship for Electronic Business at Berlin University of the Arts and is founder of the Institute of Electronic Business.

Annika Ulich is part of the research programme Data, Actors, Infrastructure at HIIG and scientific coordinator of DUCAH.

Pia Winckler studies Sustainable Development at Utrecht University and is student assistant at HIIG.

To begin with, the service could include digital but also analogue prevention tips, medical consultations and treatments. Frida would have a point of contact when she is unsure about a suddenly arising symptom she might be experiencing or if she has a question about her health in general. If a digital consultation was insufficient, she could also schedule an analogue appointment. A service like this must meet the needs of care facilities, residential partners, residents and their relatives, and caregivers. For Frida this could mean that her granddaughter needs to have the time to support her as she gets familiar with this new service and that her assisted living home would have to be able to provide the infrastructure.

USER-CENTRED DESIGN OF DIGITAL SOLUTIONS

How the elderly perceive and use technology to enable living as independently as possible is driven by their personal, social, cognitive and physical contexts. The potential user of a concierge solution needs good eyesight, good hearing, good tactile skills, supportive friends or family members to help them get used to the technology or an understanding of digital technologies. Abilities and factors like these vary greatly from person to person and change over time. Older people must not be seen as a homogeneous group of actors just because they are the same age: they will use, modify, and develop technologies in varied ways for many different and possibly contradictory reasons. While one person might be excited to use a concierge tool but cannot see the interface properly and needs voice control, Frida Müller might only need someone to encourage her to try it out. This is supported by Merkel and Kucharski (2019), who argue that it is important to include older users in all stages of the innovation and development process. Stakeholders need to carefully consider how to design, implement and adapt technologies to support aging in place. In addition to the elderly person's point of view, the perspectives and needs of their support system, their families or professional carers, need to be taken into account as well, to ensure that the implemented technology is a long-term success. There are many steps that must be taken when integrating a concierge technology to improve an older person's health and well-being. It needs to be set up, it needs to be paid for, possibly by a health insurance provider, it needs to be explained to the person using it and this person needs to see its usefulness to continue working with it. The technician installing the hardware will not be able to meet these requirements. Instead, people that are trusted by the prospective concierge user, like eldercare providers or relatives, have to be able and willing to invest time and effort. This means including their needs and priorities when developing a digital solution. Great

care must be taken to ensure that telecare systems do not feel like an imposed or even coercive measure to the person who should feel supported (Bächle, 2020). Understanding needs and individual creativity when customising systems is essential to the ethical use of digital and remote care, and this customisation process should be respected (Bächle, 2020). The development of technological products should always be viewed with one primary objective in mind: how can the elderly be supported in leading an independent, healthy life for as long as possible. This must include frameworks and policies that address people, regulation, physical spaces, relationships and participants' beliefs and attitudes.

THE LEARNING LAB AS A MULTI-DISCIPLINARY APPROACH TO RESEARCHING AND DESIGNING SOLUTIONS FOR AGING

Another aspect of aged living which has gained traction within debates of aging in the last few years is the significance of mobility and neighbourhood. The intensification of work on smart cities, transportation and digital health shows the need for forums in which exchange and mutual understanding of these objectives and logics can be facilitated and which then support sustained collaboration and co-design. Digital solutions such as smartwatch wearables with fall detection illustrate the intersection of mobility, technology and health. Such a device has the potential to encourage elderly people to move freely and feel safe, in their home but also in their local area. The Digital Urban Center for Aging and Health | DUCAH was founded as a forum to bring together stakeholders from diverse fields and to create a neighbourhood-based Learning Lab. This will serve as a central reference space for a strategic multi-disciplinary approach to research as the design of solutions for aging is gaining relevance.

The DUCAH white paper *Aging, Independent Living and Society* by Paul Jackson et al. (2021) introduces the Learning Lab. It is based on the term Living Lab but goes further. The concept of Living Labs has gained currency as a way of testing applications within an existing social environment (Keyson et al., 2017) and has been applied to areas as diverse as traffic management and cultural heritage. Furthermore, Living Labs are "user-centred, open innovation ecosystems based on a systematic user co-creation approach, integrating research and innovation processes in real life communities and settings" (European Network of Living Labs, n.d.). The Learning Lab is a European conceptualisation to facilitate the testing of innovative concepts. Within it, the adaptation of technologies and systems to people and their needs can be observed and evaluated within a live

environment. Any evaluation should take into account interoperability and collaboration in digital ecosystems as well as best practices in digitally supported care organisation and provision. Does a concierge service provide the intended assistance in a person's everyday life and if not, what should be changed? How does an elderly person like Frida Müller utilise fall detection technology and does it make her feel safer moving in her local area? Questions like these can be explored holistically in the Learning Lab with a participatory design approach and multidisciplinary teams of technologists, health care specialists, public health researchers, business scientists and collaborations with policy makers, community administrators, technology developers, care providers, entrepreneurs, finance providers and other stakeholders. Together, these individuals and organisations can work to develop and test digital health and care solutions that help preserve the existing social environment for all generations. ◆

REFERENCES

Bächle, T. C. (2020). Narrative der digitalen Überwachung. In K. Hauptmann, M. Hennig & H. Krah (Eds.), *Narrative der Überwachung. Typen, mediale Formen und Entwicklungen* (pp. 225–253). Peter Lang. https://library.oapen.org/bitstream/handle/20.500.12657/42776/9783631827475.pdf?sequence=1#page=227

Chrysikou, E., Rabnett, R., & Tziraki, C. (2016). Perspectives on the Role and Synergies of Architecture and Social and Built Environment in Enabling Active Healthy Aging. *Journal of Aging Research*, 2016. https://doi.org/10.1155/2016/6189349

European Network of Living Labs (n.d.). About Us. https://enoll.org/about-us

Jackson, P., Schildhauer, T., Ulich, A., Pohle, J., & Jansen, S. A. (2021). Aging, Independent Living and Technology. Foundation for Internet and Society. https://doi.org/10.5281/zenodo.5026032

Keyson, D.V., Morrison, G.M., Baedeker, C., & Liedtke, C. (2017). Living Labs to Accelerate Innovation. In D. Keyson, O. Guerra-Santin & D. Lockton (Eds.), *Living Labs*. Springer. https://doi.org/10.1007/978-3-319-33527-8_5

Merkel, S. & Kucharski, A. (2019). Participatory Design in Gerontechnology: a systematic literature review. *The Gerontologist, 59*(1), e16-e25. https://doi.org/10.1093/geront/gny034

HIIG FELLOWS

Wild imaginaries

Every year, our research fellows dive into a new world for *encore*. This time it's pop culture and technology. What do Lady Gaga and a text-generating machine have in common? How can Disney princesses inspire technological innovations? And how does the Terminator imagine his future with human intelligence?

SARAH CISTON

Lady Gaga, GPT-3 and Gertrude Stein have more in common than you think

News about the latest text-generating machine-learning (ML) tools like GPT-3 usually focuses on their scary, human-like accuracy or the ever-expanding scale of their massive datasets. You'd rarely mention them when thinking of your favourite karaoke songs or Modernist poets – but perhaps you should.

While they have much in common with hit pop songs and avant-garde poetry, GPT-3 and similar transformer models in natural language processing (NLP) are not the panacea they've been called. "Writers, as they have always been up to now, are writing machines", said Italo Calvino in 1987. As computational machine writing grows popular, let's remember that machine understanding does not mean comprehension in the same vein as our own ways of knowing. Instead, NLPs find patterns. They sift through the truly massive amounts of language that has been (depending on the task) recorded in audio or scraped from the web – every stupid thing someone wrote on Reddit is probably in there somewhere – and they try to emulate so-called natural language from these millions of examples, noticing which sounds and letters most often follow which, rather than what and why words actually mean what they mean.

As such, GPT-3 works like Lady Gaga, it works like Gertrude Stein and it sometimes sounds like both of them, too. Let me know if you can tell which wrote each of these:

1 Don't be, don't be, don't be. Give yourself prudence and love your friends.

2 A single image is not splendour.

3 You can say things like, “It’s true”, “It’s true”, “It’s true”, or “It’s true”, You can say things like, “I’m doing this on my mobile”

4 A willow and no window, a wide place stranger, a wideness makes an active centre.

5 I’m worried about who I am in the world. I’m worried about who I am in myself. I’m worried about who I am in the world. I’m worried about who I am in myself.

6 your psycho, your vertigo schtick, want you in my rear window

How did you do? Notice the attention each gives to repetition and sound: consonance, assonance, rhythm. Each of the selections seem to have some meaning, yet, upon analysis, it unravels into word salad. Our minds do most of the work through the frisson of novel words near one another. For generations, that has been part of the pleasure of poetry – and it can be part of the joy of NLP too, if we are willing to stretch beyond the narrow (impossible) task of what we call optimisation.

We have been unwittingly adapting ourselves and our modes of language to be better understood by machines. *Linguistic capitalism* turns text into profit through search, autocomplete and predictive advertising algorithms—all of which standardise our language into things we can shop for. But attention to errors in ML tools reclaims space for other kinds of value, such as poetry. So, when Journey sings, “streetlight people,” or when Prince sings, “even doves have pride,” think of GPT-3 spitting out the future of avant-garde poetry for your next karaoke party.

1. Lady Gaga. 2. Gertrude Stein 3. GPT-3 4. Gertrude Stein 5. GPT-3 6. Lady Gaga

LUBNA RASHID

How Jasmine, Belle and Rapunzel can inspire tech innovation in times of crisis

The sultan's rebellious daughter, the mischievous long-haired princess and the adventurous bookworm certainly have more to teach us than "fall in love with a good guy and live happily ever after". As a little girl growing up in a conservative, conflicted part of the world, my fairy tale addiction might have played a role in my current view of the world and career choices. Let me tell you why.

All three girls were extremely intelligent yet trapped in captivity: a guarded palace, a lonely tower or a beast's castle. What better symbolises the entrapment that today's motivated, ambitious girls feel in contexts that clip their wings and push them into traditional roles? Nonetheless, those three imaginative girls managed to mobilise the limited resources in their surroundings to creatively solve problems, repurpose those resources, and recombine them to optimise desired outcomes like remarkably resilient bricoleurs. Clearly, they have enough proactiveness, innovativeness and risk-taking capacity to rank them high up the entrepreneurial orientation scales.

But what does this have to do with tech entrepreneurship specifically? Well, remember what kinds of resources they mainly relied on? A flying carpet, talking tea pots and hair with a mind of its own are all intelligent non-living objects, or *artificial intelligences*, that greatly fostered their entrepreneurialism. Replace those with smartphones, laptops, open source code and e-learning resources, and maybe some real-life magic would unfold. Clearly, no amount of princess stories or tech gadgets would fix the broken systems and human rights violations that crush away millions of girls' dreams every day, but who knows what creative sparks this could ignite. Perhaps entrepreneurship education, in my home country and beyond, can start with a fairy tale and an internet connection.

MAURICE JONES

Imagining humanity: the Terminator's vision of a future together with human intelligence

It is 1984. The Cyberdyne Systems Model 101 (T-101) also known as the Terminator just became one of the greatest and scariest villains in sci-fi history. Upon his return in the 1991 sequel *Terminator 2: Judgement Day,* the T-101 was equally destructive, but he had undergone a radical inner transformation from villain to protector, martyr and near father figure. My research usually investigates how we humans imagine our future together with artificial intelligence. To understand the Terminator's change of robot heart, we need to flip the coin and explore this android's vision of his future together with human intelligence.

After a careful discourse analysis, which consisted of binge-watching these two movies back-to-back, I began to garner a clear picture that is best described through two essential quotes. As the Terminator itself is an android of few words, I will refer to quotations from one of the supporting characters. In the 1984 movie, the character Kyle Reese refers to the Terminator:

"Listen and understand. That Terminator is out there. It can't be bargained with. It can't be reasoned with. It doesn't feel pity, or remorse, or fear. And it absolutely will not stop, ever, until you are dead."

This quote conveys a clear sense of how the Terminator imagines its future together with humanity – it sees none, because humanity needs to be terminated. In contrast in the sequel, one encounters the following dialogue:

John Connor: "You're not here to kill me. I figured that out for myself. So, what's the deal?"

Terminator: "I'm here to protect you."

Contrasting these two quotes gives an unquestionable indication that the Terminator's future vision changed between the two movies – from a future that entailed terminating humanity to a more hopeful one of coexistence with humanity. A future that is worth protecting.

There is a lot to learn here from the Terminator. For a murderous killer machine that was set out to destroy humanity to not only imagine a different future alongside human intelligence but to actively fight for its creation surely shows the power of artificial imagination. Following in T-101's footsteps, we might be advised to step down from our high horses and start imagining, building and protecting our desirable futures together with artificial intelligence instead of projecting our dystopian visions of the future on the unfortunate algorithms, androids and other artificial beings. The ironic part now is that the change of robot heart was indeed programmed by a future John Connor in order for the Terminator to travel back in time as a protector, a luxury we most likely will not have in our technological future. ◆

Your response to the question: How does your job relate to the internet?

KARINA PREISS

Hats off! Have you ever wondered who holds all the threads together at the institute? Karina Preiß was responsible for the establishment process of the Alexander von Humboldt Institute for Internet and Society, Germany's first internet research institute. She obtained her doctorate in Organisational Structures in Virtual Organisations, and has used her knowledge over the last ten years to excel at the virtual organisaton of real structures at HIIG – as the institute's Managing Director – congratulations on the anniversary!

How do you react to a research proposal being rejected?

How do you react to a research proposal being accepted?

How do you react when your colleagues come up with overblown research ideas?

What is your favourite activity with your colleagues?

FROM A SINGLE FORMAT TO AN ADAPTABLE SUPPORT SYSTEM

DIGITAL TRANSFORMATION IN SMALL AND MEDIUM-SIZED ENTERPRISES

Small and medium-sized enterprises (SMEs) tend to have a rather conservative attitude toward the risk of developing innovations. Their digitalisation activities are often reactive and selective, reflecting their typically narrow range of products and services and their limited financial and human resources. Strategic innovation activities are put on hold in favour of short-term projects, and there is too little cooperation with external partners to tap into new knowledge and new business areas.

1621
SMEs participated in

24 Workshops
39 AI Consultations
17 Strategic sprints
8 Startup Tours

Our work over the past five years has shown that SMEs require tailored support to tackle the challenges of digitalisation and disruptive technologies. The following qualities ensure success and long-term effectiveness:

Diversity and adaptability: Only diverse access and asynchronous offerings can address the heterogeneity of SMEs. Formats pre-sorted by digital maturity level serve to adapt to concrete needs and prior knowledge.

Applicability and tangibility of technologies: In particular, technologies such as AI, which appear to be knowledge- and capital-intensive, must be made usable and tangible for SMEs. Consultation formats give the opportunity to think about concrete use cases in SMEs and thus help to reduce barriers to innovation.

Building resources: In particular, cultural and human resources are needed to shape the digital transformation from within SMEs. The teaching of digital soft skills enables organisations to proactively and holistically harness new technologies for themselves and to critically assess and deploy them.

Needs assessment and accompanying research: Close dovetailing of practice-oriented support and research is necessary in order to promote and guide the digital transformation in the sector in a long-term, sustainable and needs-oriented manner.

Holistic and sustainable support systems must be developed with and for the target groups in order to be effective. Although German SMEs still have a great deal of potential in terms of digitalisation, they also have unique resources that enable them to actively and innovatively shape digitalisation.

The Federal Ministry of Economic Affairs and Energy-funded programme *_Gemeinsam Digital* supports SMEs by offering expert knowledge, demonstration centres and best-practice examples, among other things. The Alexander von Humboldt Institute for Internet and Society (HIIG) was part of the project between 2016 and 2021.

5 Years

31 Researchers

20 Video tutorials

15 Podcasts

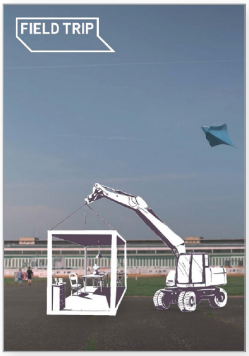
16 Case studies

18 Blog posts

LONGREADS 2021

Dissertations and books published by HIIG researchers

DISSERTATIONS



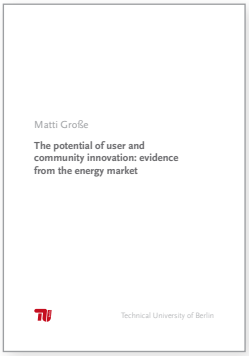
Frédéric Dubois

Interactive Documentary Production and Societal Impact: The Case of Field Trip

Film University Babelsberg KONRAD WOLF

The thesis is based on a research-creation approach which includes an empirical case study of the interactive documentary (i-doc) Field Trip (2019). By articulating a societal impact framework, it contributes to a better understanding of the cultural value of contemporary interactive storytelling practice.

fieldtrip.berlin



Matti Große

The potential of user and community innovation: evidence from the energy market

Technical University of Berlin

This dissertation develops typologies and frameworks that make the diversity of user innovation activities clearly measurable and thus comparable. It analyses the conditions under which user innovation is triggered, the extent to which user-innovators are active and to what extent their ideas support societal transformation processes.

EDITED VOLUMES

Thomas Döbler, Christian Pentzold, Christian Katzenbach (Eds.)
Räume digitaler Kommunikation. Lokalität – Imagination – Virtualisierung
Herbert von Halem Verlag · ISBN 978-3-86962-440-2

This volume contributes to the conceptual and theoretical understanding of spaces of digital communication. By means of empirical work and case studies, the authors show processes as well as structures influencing communication practice.

Jörg Pohle, Klaus Lenk (Eds.)
Der Weg in die „Digitalisierung" der Gesellschaft. Was können wir aus der Geschichte der Informatik lernen?
Metropolis Verlag · ISBN 978-3-7316-1461-6

Allegedly, the misleadingly named "digitalisation" is revolutionising all areas of our connected lives. The authors want to draw lessons for the future from the disappointments, but also the undeniable successes of information technology. To what extent is past computer science thought still helpful today in drawing a clear picture of the informatisation of society and its implications?


E

What is your favourite place at the institute?

encore

Volume 2021/2022
Published January 2022

PUBLISHER

Alexander von Humboldt Institute for Internet and Society
Französische Str. 9
10117 Berlin

EDITORIAL TEAM

Hadi Asghari, Nadine Birner, Stephan Bohn, Nicole Butters,
Jan Sebastian Götte, Lena Marie Henkes, Sonja Köhne,
Nastasja Krohe, Katharina Mosene, Björn Scheuermann,
Jessica Sereflioglu, Moritz Timm, Katrin Werner, Larissa Wunderlich

DESIGN

Larissa Wunderlich

CREDITS

Photo p. 18 by Princeton University / Tor Birk Trads
Photo p. 39 by IFSH
Photo p. 63 by Paige Zangoglia
Illustrations pp. 115/116/119 by Fehmi Baumbach
Photos pp. 90, 174/175 and 182 by
Mathias Völzke Photography

Print ISBN 978-3-9821760-7-9
Digital ISBN 978-3-9821760-8-6

Alexander von Humboldt
INSTITUTE FOR INTERNET
AND SOCIETY

Alexander von Humboldt
INSTITUTE FOR INTERNET
AND SOCIETY