encore

MAGAZINE ON INTERNET AND SOCIETY

VOLUME 2025/2026

hiig.de/encore2025

SELECTED CONTENT ON OUR FIVE RESEARCH TOPICS IN FOCUS

Platform governance

Artificial intelligence and society

Science in times of polarised public issues

Digital organising and the future of work

Digitalisation and sustainability

EDITORIAL

2025 marked a breakthrough year for the Alexander von Humboldt Institute for Internet and Society (HIIG). The Joachim Herz Foundation, the Stiftung Mercator and the Foundation Science and Democracy have committed to supporting our work in the years ahead. With new wind filling our sails, we are embarking on a bold new chapter, reshaping our research agenda to meet the urgent demands of digitalisation.

However, this new momentum is not the only reason to celebrate! This issue of encore highlights our research's impact beyond academia – we examine platform governance, uncover invisible content moderation labour, assess the social effects of AI, explore the realities of digital work, investigate sustainability in data-driven technologies and analyse science's role in polarised debates.

Our approach to scientific impact is a processual one. It entails the ongoing exchange and interaction between research and society, especially with those charting the course of digital transformation. As you will see, this includes collaborating with municipalities to tackle digitalisation backlogs, discussing sustainable use of AI in classrooms and engaging with European regulators on fair content moderation.

We hope you enjoy reading!

Björn Scheuermann,
Director at HIIG

Freia Kuper,
Researcher at HIIG

Katharina Mosene,
Researcher at HIIG

Platform governance

PI

AI

Artificial intelligence and society

FOCUS

Platform governance

Platform governance refers to the rules, regulations and frameworks for managing digital platforms in our society. This includes social networks, online services, digital marketplaces or messaging services, to name a few. These digital platform ecosystems facilitate the exchange of information, goods and services in our daily lives. As an integral part of our modern communication, they also have a profound impact on public discourse. In our research on sustainable platform governance, we therefore examine how economic goals, individual rights and societal values can be aligned or at least harmonised in these online communication spaces. This includes questions related to online platform regulation, competition law, freedom of speech, individual autonomy and (democratically anchored) decision-making.

VISIT TOPIC OVERVIEW ONLINE

JELLA OHNESORGE

Counting without accountability? An analysis of the DSA's transparency reports

The Digital Services Act aims to hold platforms more accountable for illegal content by demanding greater transparency. Platforms like Facebook, Instagram, Tiktok or X must now publish detailed reports, showing, for example, how many posts they removed and how quickly. These reports are meant to give regulators, researchers and the public insight into how well platforms are enforcing the rules. But do the reports really deliver what they promise? Or is this measure just a new but ultimately useless addition to the flood of EU reports without any actual improvements in practice?

Online platforms like Instagram, TikTok or X have become an integral part of everyday life. However, despite their societal relevance, these privately owned platform companies remain largely opaque when it comes to understanding how they work. They rarely explain how they choose which content to distribute, remove or suppress, even though their decisions determine what we see online.

HOLDING PLATFORMS ACCOUNTABLE

One of the core aims of the new EU digital rulebook, including the Digital Services Act (DSA), is to regulate how platforms handle illegal content and enable more effective action against it. What constitutes illegal content is not explicitly codified in the DSA but rather ultimately depends on what is illegal under Union or Member State law. This often includes child sexual abuse material, incitement to terrorism, illegal hate speech or infringement of intellectual property rights.

Platforms are to be held accountable for their reactions once they are made aware of illegal content. For example, if a user reports a post or video that they believe contains illegal hate speech, the platform must review the report and decide on an action such as deleting or restricting the post. Later, the platform must also disclose how many such cases were handled within a given timeframe and what they did about the reported content in a transparency report.

One of the DSA's main aims is to increase the accountability of platforms by promoting transparency. A key assumption underlying the DSA is

that digital services may pose systemic risks to society. Examples of such risks include widespread disinformation and the undermining of electoral integrity. To identify and limit these risks at an early stage, maximum transparency is necessary. This is supposed to enable public authorities, researchers and civil society to recognise potential systemic risks, and allow individual users to understand and assert their rights.

THE DSA'S TRANSPARENCY REPORTS

Transparency reports are a key accountability measure under the DSA. Under Articles 15, 24 and 42, platforms must publish comprehensible reports on their content moderation activities. These reports must include information on the types of illegal content moderation and the actions taken. Content moderation actions can include deleting the content, demoting it or geo-blocking it in a specific country. The reports must be publicly available in a machine-readable format (Art. 15(1) DSA). Most platforms provide them as PDF documents or HTML pages on their websites. They are also collected and linked on an EU website (European Commission, 2025).

But how much transparency do these reports actually provide? Are they really suited to uncover how platforms decide what to take down and what not to? And most importantly, are they an adequate measure to increase platform accountability? By analysing the transparency reports of selected online platforms, I argue that current transparency reports fall short of delivering true accountability with regard to the moderation of illegal content. Even though a new standardised reporting template has recently been introduced,

and many hope it will improve the situation, I argue that the template can only address some of the inadequacies of the current reports while potentially creating new problems.

TRANSPARENCY REPORTS IN PRACTICE: THREE OBSERVATIONS

What do the transparency reports actually reveal about how platforms handle illegal content? Based on a qualitative analysis of the two reporting rounds of DSA transparency reports published in 2024 by seven very large online platforms (VLOPs, platforms with more than 45 million monthly average users), namely Instagram, Facebook, LinkedIn, Pinterest, Snapchat, TikTok and X, I examined how these VLOPs fulfill their transparency requirements.

Although the European Commission (EC) provides guidance on the content of transparency reports, it does not specify their structure or level of detail. The general idea seems to have been to give the platforms some leeway, on the one hand, and on the other, to see whether it would be possible to build on best practices and refine the specifications later.

The new template is intended to clarify the expected form, content and level of detail in the reports (European Commission, 2024). However, the regulation has only been in force since July 2025 and has not yet been applied in practice. In the available reports, each platform has interpreted the DSA specifications independently. My analysis revealed three key findings that highlight the variety of approaches used by platforms for their transparency reporting obligations and the limitations of the current reporting format.

OBSERVATION 1: DISCONNECTED DATA POINTS

One major issue is the lack of connection between different data points within the individual reports. For example, figures relating to moderation decisions, user complaints or automatic deletions are often only presented as individual values, with no reference to one another. This means that it is hard to establish a relationship between data, which makes it harder for researchers to interpret or integrate them into meaningful analyses.

An example of this is the reporting of authority orders under Article 9 of the DSA, which sets out how Member State authorities, such as national courts or Digital Services Coordinators (in Germany, the Bundesnetzagentur), can request that platforms take action against illegal content. The term "order" can be misleading though, as platforms are not required to delete content that is referred to them by a Member State authority. Instead, they review the content independently and then decide whether to take action.

In its 2024 transparency reports, Facebook provided two separate tables: one lists the number of orders received per Member State, the other the number of orders by content type, such as terrorist content, illegal speech, etc. (Facebook, 2024, p.3–5; Facebook, 2025, p.4–6). However, these two tables are not linked, which makes it impossible to find out, for example, how many orders to act against terrorist activity were issued by Italy against Facebook. This lack of cross-referencing severely limits the analytical value of the data.

OBSERVATION 2: ARBITRARY AND INCONSISTENT CATEGORIES

There is no standardised categorisation of illegal content across platforms, sometimes not even within the individual reports. Each platform has created its own set of categories, loosely based on EU or Member State law, but ultimately inconsistent and often arbitrary.

LinkedIn, for example, uses the label "Illegal or harmful speech" (LinkedIn, 2024, p.17) while Facebook uses terms like "Hate speech" and "Misinformation" (Facebook, 2024, p.4). Pinterest takes a different approach and directly refers to the specific laws that a piece of content is said to violate.

Most platforms also include a vague catch-all category such as "Other illegal content". For example, Instagram's transparency report for April to September 2024 states that it received 91 orders from authorities to act against "other" types of illegal content, accounting for almost one third of all cases (Instagram, 2024, p.4). Facebook received 113,638 user notices for "other illegal content", accounting for around 45% of the total 248,748 notices received.

Furthermore, the fact that some platforms use different categories for authority orders and user-submitted reports of illegal content adds unnecessary complexity. This makes it difficult to compare between platforms and further complicates the picture of how illegal content is handled globally.

OBSERVATION 3: OPAQUE DECISION-MAKING

The most striking issue, perhaps, is the lack of clarity surrounding what happens after a platform receives a user notice or an authority order. In many reports, it is unclear what action was taken, on what basis, and whether any action was taken at all. While most platforms report how many orders they have received from authorities, they do not report whether they responded by deleting the reported post, deleting the account, geo-blocking the content in a specific country or reducing a post's visibility.

LinkedIn merely hints whether "at least some action was taken" (LinkedIn, 2025, p.16–17) in response to an authority order, but does not provide further details. Pinterest is the notable exception: it clearly indicates whether it deactivated content, restricted it geographically or limited its distribution (Pinterest, 2025).

Another issue is the considerable confusion surrounding the two types of reporting mechanisms: the specific mechanism for reporting illegal content under Article 16 of the DSA, and the general channels that platforms have in place for reporting any type of rule violation (e.g. content that violates a platform's advertising policy but not any laws). Article 16 requires platforms to have a reporting mechanism through which users can report potentially illegal content in a precise and substantiated way. For example, if a user believes that a post incites terrorism, they must be able to report it to the platform in a way that clearly indicates the potential illegality of the content.

In practice, all user reports, regardless of the reason given for the report, are first reviewed for violations of the platforms' own rules, such as community guidelines or advertising policies. Therefore, at the outset of the review process, it is irrelevant whether it is an Article 16 notice or a different kind of user report. If an Art. 16 notice is found to violate a platform rule and if this leads to the reported content being deleted globally, it is never checked to see if it had also violated the law – even if originally reported for that reason. So, if the terrorism-inciting post in our example also violated a platform's advertising policy, it would never be reviewed for breaching any anti-terrorism laws. This means: the content would disappear globally, not just in the country where it might be unlawful.

While this approach is clearly efficient – Why bother blocking a post in only one country if it violates a platform rule and would be deleted globally in any case? – it raises questions about who decides how public debate takes place and on the basis of which rules.

Neither LinkedIn nor Snapchat explicitly distinguish between actions taken following a user report based on the law and those based on internal policies. Snapchat even argues that breaches of the law are automatically covered by their own rules, as a violation of their Community Guidelines includes "reasons of illegality" (Snapchat, 2024). This seems to be in tension with Article 15 of the DSA, which clearly states that providers must specify whether an action was taken on the basis of the law or their own terms and conditions.

In summary, the reports analysed here reveal major inconsistencies and blind spots. From disconnected data points and arbitrary categories to the opaque reasoning behind content moderation decisions, the reports currently fall short of offering real transparency. Against this backdrop, the following section outlines three key criticisms of the current reporting system and considers whether the new template might address some of these shortcomings.

THREE KEY FLAWS IN PLATFORM TRANSPARENCY REPORTS

FLAW 1: THE PROVIDED DATA IS BORDERLINE UNUSABLE.

The big differences in amount, level of detail, operationalisation and presentation of data across reports make it quite hard to compare the platforms and assess what measures are most effective in combating illegal content. In many cases, it is not even possible to establish connections between data points within individual reports, which further hinders any meaningful evaluation. The new template, which is an Excel spreadsheet where platforms can enter their data, should help to address some of these problems. For one thing, comparability is likely to improve if all platforms provide their information in the same format. The template also introduces fixed categories of illegal content and requires that the category "other" be described. However, the template only asks platforms to report the "number of items moderated" (European Commission, 2025), without specifying what type of content moderation action was taken. This is surprising, given

that Article 15 of the DSA requires platforms to report how many notices they received under Article 16, and to categorise these by "any action taken pursuant to the notices" (Art. 15(1b) DSA). The new template, however, does not seem to include this requirement.

FLAW 2: THE PROCESS OF MODERATING ILLEGAL CONTENT REMAINS LARGELY OPAQUE.

Even when report data does show which actions came in response to which notices or reports, the underlying process remains a black box. Questions such as "What criteria are used to decide cases?" and "What legal expertise do content moderators have?" remain unanswered. Snapchat, for example, has received 82,011 Art. 16 notices for content that potentially violates rules against false information during the reporting period between January and June 2024. Out of these, 106 pieces of content were deleted, 255 accounts were issued warnings and 12 accounts were locked. Setting aside the absurdly low number of actual actions, it is impossible for us to know why Snapchat deleted content in some cases and not in others, or why it locked accounts in some cases and merely issued warnings in others. The new template does not ask for that kind of information. So, it is unlikely that we will see an improvement in this regard.

FLAW 3: PLATFORM RULES REMAIN THE GOLD STANDARD.

While it makes sense for efficiency reasons and is completely in line with the DSA, platforms still primarily check if content violates their own rules before

or instead of focusing on EU or Member State law. However, this renders the separate mechanism for reporting illegal content frankly obsolete. It also raises questions about which rules are considered more important: those set by a private platform company or democratically legitimised laws. The new template cannot fundamentally challenge this hierarchy, nor does it make more visible the rules and criteria that platforms use to make content moderation decisions. Instead, the template further entrenches the already limited amount of information that platforms make available.

THE TEMPLATE IS NOT THE SAVIOUR SOME MAKE IT OUT TO BE

The analysis of the 2024 Transparency Reports from seven very large online platforms (VLOPs) (Instagram, Facebook, LinkedIn, Pinterest, Snapchat, TikTok and X) reveals significant shortcomings in the way these companies report on the moderation of illegal content. Key data points are not connected, categories of illegal content are applied inconsistently and arbitrarily, and the reasoning behind content moderation remains largely opaque.

These gaps make it difficult to assess how platforms actually respond to illegal content and thus to measure the adequacy and effectiveness of their response. Still, the newly introduced EU template for transparency reports is a step towards greater comparability and clarity as it standardises reporting formats and categories. It thereby may help reduce the inconsistency observed so far. However, the template leaves important blind spots unaddressed. Most notably, it does not require platforms to explain their reasoning behind moderation decisions or distinguish clearly between enforcement based

on law versus internal rules. There might also be potential issues with platforms adhering only to the template's minimum requirements, which could reinforce existing shortcomings and further limit the availability of meaningful information. Thus, neither the transparency reports nor the new template currently achieve accountability through transparency.

EU REGULATIONS AND TEMPLATES

European Commission (2025). Annex I – Transparency reports template [Microsoft Excel file]. *Publications Office of the European Union*.

European Commission (2024). *Commission Implementing Regulation (EU) laying down templates concerning the transparency reporting obligations of providers of intermediary services and of providers of online platforms under Regulation (EU) 2022/2065 of the European Parliament and the Council*. https://eur-lex.europa.eu/eli/reg_impl/2024/2835/oj/eng

European Commission (2022). *Regulation (EU) 2022/2065 of the European Parliament and of the European Council of 19 October 2022 on a Single Market For Digital Services and Amending Directive 2000/31/EC (DSA)*. https://eur-lex.europa.eu/eli/reg/2022/2065/oj

TRANSPARENCY REPORTS

European Commission (2025). How the Digital Services Act enhances transparency online. Retrieved 20. September 2025, from https://digital-strategy.ec.europa.eu/en/policies/dsa-brings-transparency

Facebook (2024). Regulation (EU) 2022/2065 Digital Services Act Transparency Report for Facebook. April–September 2024. *Facebook*. https://transparency.meta.com/reports/regulatory-transparency-reports

Facebook (2025). Regulation (EU) 2022/2065 Digital Services Act Transparency Report for Facebook. October–December 2024. *Facebook*. https://transparency.meta.com/reports/regulatory-transparency-reports

Instagram (2024). Regulation (EU) 2022/2065 Digital Services Act Transparency Report for Instagram. April–September 2024. *Instagram*. https://transparency.meta.com/reports/regulatory-transparency-reports

Instagram (2025). Regulation (EU) 2022/2065 Digital Services Act Transparency Report for Instagram. October–December 2024. *Instagram*. https://transparency.meta.com/reports/regulatory-transparency-reports

LinkedIn (2024). Digital Services Act Transparency Report. January–June 2024. *LinkedIn*. Retrieved 22.10.2025, from https://www.linkedin.com/help/linkedin/answer/a1678508

LinkedIn (2025). Digital Services Act Transparency Report. July–December 2024. *LinkedIn*. Retrieved 22.10.2025, from https://content.linkedin.com/content/dam/help/tns/en/February-2025-DSA-Transparency-Report.pdf

Pinterest (2024). Digital Services Act Transparency Report. January–June 2024. *Pinterest*. Retrieved 22.10.2025, from https://policy.pinterest.com/en/transparency-report-h1-2024

Pinterest (2025). Digital Services Act Transparency Report. July–December 2024. *Pinterest*. Retrieved 22.10.2025, from https://policy.pinterest.com/en/digital-services-act-transparency-report-jul-2024-dec-2024

Snapchat (2024). European Union Transparency | Snapchat Transparency. January–June 2024. *Snapchat*. Retrieved 22.10.2025, from https://values.snap.com/privacy/transparency/european-union

TikTok (2024). TikTok's DSA Transparency report. January–June 2024. *TikTok*. Retrieved 22.10.2025, from https://sf16-va.tiktokcdn.com/obj/eden-va2/zayvwlY_fjulyhwzuhy[/ljhwZthlaukjlkulzlp/DSA_H2_2024/TikTok-DSA-Transparency-Report-Jan-to-Jun-2024.pdf

TikTok (2025). TikTok's DSA Transparency report. July–December 2024. *TikTok*. Retrieved 22.10.2025, from https://sf16-va.tiktokcdn.com/obj/eden-va2/zayvwlY_fjulyhwzuhy[/ljhwZthlaukjlkulzlp/DSA_H2_2024/Corrected%20Data/TikTok%20-%20DSA%20Transparency%20report%20-%20July%20-%20December%202024%20-21.03.2025.pdf

X (2024). DSA Transparency Report. April–September 2024. *X*. Retrieved 22.10.2025, from https://transparency.x.com/dsa-transparency-report.html

X (2025). DSA Transparency Report. October 2024–March 2025. *X*. Retrieved 22.10.2025, from https://transparency.x.com/assets/dsa/transparency-report/dsa-transparency-report-april-2025.pdf

ANN-KATHRIN WATOLLA

Debunking assumptions about disinformation: Rethinking what we think we know

Disinformation has become one of the buzzwords of our time. The term appears so frequently in headlines and political warnings that it seems to be thoroughly understood. Given the amount of attention it receives, it is easy to assume that we already grasp the phenomenon. But do we? Reliable evidence remains quite limited. In this article, Ann-Kathrin Watolla uses a recent review to challenge three common assumptions about disinformation, examining what we do and don't know.

READ FULL ARTICLE

ALINA KONTAREVA

Beyond Big Tech: National strategies for platform alternatives

Global platforms such as Google, Amazon and YouTube have become an essential part of digital infrastructure supporting the internet as we know it. However, their dominance raises concerns about dependencies and lock-in effects. While regulation has long been the main response, some states have opted for a more proactive path: they've built their own platform alternatives in order to gain control over critical digital services and data. In this article, Alina Kontareva analyses the motivations behind these strategies in China, Russia and India. What lessons could Europe learn from this as it seeks to shape its own digital environment?

READ FULL ARTICLE

MAURICE STENZEL, KATHARINA MOSENE
& FREDERIK EFFERENN

Inside content moderation: Humans, machines and invisible work

Content moderation relies on platform rules, technical systems and legal requirements, yet its consequences reach far beyond removing harmful posts. The division of labour between automated systems and human moderators has repeatedly reached its limits. Much of this work is outsourced to countries such as the Philippines or Kenya, where people review distressing content under precarious conditions. The authors of this article examine how these practices and guidelines shaped in the Global North reproduce or even amplify inequalities – for instance along lines of gender, origin or ethnicity.

READ FULL ARTICLE

INTERNET POLICY REVIEW · SPECIAL ISSUE

Content moderation on digital platforms: Beyond states and firms

New regulatory efforts in Europe and recent shifts in platform policy – such as Meta's decision to end fact-checking partnerships – highlight the growing influence of non-state actors in shaping digital governance. However, how fair, inclusive and effective is their role in content regulation? This special issue, edited by Romain Badouard and Anne Bellon, brings together twelve contributions that trace the diverse roles of civil society actors. These range from NGOs and journalists to designers and researchers, and unfold across a variety of institutional and symbolic contexts.

EXPLORE SPECIAL ISSUE

ZINE ON PLATFORM REGULATION

What does the Digital Services Act actually do?

The Digital Services Act (DSA), created by the European Union, is the first uniform legal framework of its kind to create a safer and more transparent online environment. Print out this easy-to-read mini magazine by Ann-Kathrin Watolla and Jella Ohnesorge and delve into a straightforward introduction to the obligations the regulation imposes on digital platforms. What steps can be taken to ensure that fundamental rights, such as freedom of expression and non-discrimination, are better protected?

START FOLDING

SPOTLIGHT ON CURRENT ISSUES

Disinformation and the German election

In our new HIIG Spotlight online format, we explore current socio-political issues through short contributions and engaging discussions. In February, we took the opportunity of the German federal election to discuss Meta's withdrawal from fact-checking and Elon Musk's support for far-right groups. Meanwhile, the EU is introducing the Digital Services Act, a legal framework that reshapes platform accountability. How do these changes overlap, and what will their long-term impact be on democracy, far beyond the election?

WATCH FULL SPOTLIGHT

STRENGTHENING TRUST

This Code of Conduct offers ten guidelines for fair, transparent and accountable automated and human decision-making in content moderation. While machines filter harmful material and humans review sensitive cases, this interaction has its limits. But our code addresses risks like discrimination and inadequate oversight, offering sector-specific insights to complement the Digital Services Act.

EXPLORE CODE OF CONDUCT

IMPACT SPOTLIGHT
2025

We provide platforms and regulators with a practical, research-based framework for building fair and transparent online spaces.

The Human in the Loop? project examined the semi-automated processes involved in content moderation on digital platforms. The project's findings were translated into a code of conduct comprising ten guidelines on human-machine decision-making that address challenges such as discrimination, oversight and the well-being of human moderators.

In November, these guidelines were presented to digital services coordinators from 13 EU Member States, as well as representatives of the European Commission, at an official event. Interest was keen among participants; they welcomed the code as a robust, research-grounded framework to inform future regulatory decisions by complimenting the Digital Services Act with practical, sector-specific standards and concrete implementation measures.

Artificial intelligence and society

AI

PI

Platform governance

Sc

Science in times of polarised public issues

FOCUS

Artificial intelligence and society

Artificial intelligence unveils a world where the capabilities of technical systems aim to be similar to those of humans. But, AI isn't just about algorithms; it's deeply interwoven with our society. The future of AI technologies is strongly interlinked with the automation of social processes and will touch every facet of our lives: from tailoring your social media feeds to playing a part for innovations in healthcare and even climate research. It's beyond just the screens we scroll; it's in our offices, our hospitals, our roads and even in the cutting-edge robotic systems we design. Our research investigates the interplay of AI within the political, social and cultural landscapes, and explores the impact of AI use and AI discourses on society.

VISIT TOPIC OVERVIEW ONLINE

KATHARINA MOSENE

Unwillingly naked: How deepfake pornography intensifies sexualised violence against women

A seemingly innocent holiday photo becomes the template for a highly realistic nude image – generated by artificial intelligence (AI), circulated online, without the subject's knowledge or consent. What sounds like science fiction is already a disturbing reality: tens of thousands of so-called deepfake pornographic images are created daily using freely accessible AI tools. These are computer-generated images or videos that simulate nudity or sexual acts. Women are disproportionately affected. This is no coincidence: sexualised violence is deeply rooted in society and digital technologies are significantly amplifying this reality. This article explores how deepfake pornography digitally perpetuates existing structures of violence – and what must be done to offer better protection for those affected.

Deepfake pornography is not an isolated phenomenon. It is a particularly insidious form of image-based sexual violence (bff, 2024). This refers to digital assaults that use visual material to humiliate individuals or violate their sexual autonomy. Examples include upskirting (Polizei Nordrhein-Westfalen, 2021) – secretly photographing under skirts – and the non-consensual distribution of intimate images, often euphemistically referred to as "revenge porn" (Frauen gegen Gewalt e.V., 2025), or as in this case, the dissemination of fake, AI-generated depictions of nudity or intimacy. These images are frequently distributed anonymously via messaging services, imageboards (Wikipedia, 2025) or pornographic platforms. Anyone can fall victim to deepfake pornography.

The psychological toll these attacks take becomes evident in the voices of survivors. Danielle Citron, legal scholar and professor at the University of Maryland, describes deepfakes as an "invasion of sexual privacy". In an interview with Vice magazine, she quotes a survivor saying: "Yes, it isn't your actual vagina, but [...] others think that they are seeing you naked." Citron continues: "As a deepfake victim said to me – it felt like thousands saw her naked. She felt her body wasn't her own anymore" (Cole, S., 2019).

A CLICK AWAY FROM HARM: THE EASE OF CREATING DEEPFAKES

But how are such deepfakes created—and who is behind them? What once required technical expertise, time and powerful computers is now (perhaps too) easily within reach. Deepfake images and videos can be generated using a smartphone and a single social media photo. So-called nudifier apps

and browser-based services openly offer their tools online: users upload any image and, within seconds, the person pictured appears undressed (Cole, S., 2019). The first results are often free, followed by an offer for paid subscription.

These are far from isolated cases. An investigation by *netzpolitik.org* revealed a wide range of providers generating thousands of such images daily. The business is booming. A study by 404 Media further illustrates the scale: Numerous AI-powered video generators – particularly from Chinese companies – offer minimal safeguards against the production of non-consensual pornographic content (Maiberg, E. 2025). These tools are already being used en masse to create disturbingly realistic sexualised deepfake videos, using nothing more than a portrait photo. Once uploaded, these videos circulate in dedicated online communities and are nearly impossible to remove.

FROM TAYLOR SWIFT TO SCHOOLGIRLS: A SHIFTING TARGET GROUP

What makes the issue especially concerning is that most deepfakes feature bodies read as female. One reason lies in the training data behind the systems: many AI models were trained on millions of images of naked women. The result is a structurally biased technology that doesn't merely replicate gender-based violence – it amplifies it. What emerges is a deeply gendered, automated form of digital violence – primarily targeting women (SecurityHero, 2023).

Initially, it was public figures who were targeted: actresses, influencers, female politicians (Ajder, H. et al., 2019). But as the tools became more accessible, the target group shifted. Today, it is often girls and women from users' immediate social environments who are affected: classmates, colleagues, neighbours. In Spain, for example, AI-generated nude images of schoolgirls circulated in messaging groups caused a scandal already in 2023 (Köver, C., 2023). In Pennsylvania, a teenager was arrested the following year for creating deepfake nudes of his female classmates (Der Standard, 2024).

The full extent of the harm remains largely hidden. Reliable data are scarce. Many victims are not even aware that manipulated images of them are being circulated online.

A SYSTEMIC FORM OF INTERSECTIONAL VIOLENCE

This particular form of digital abuse is systemic. As legal scholar Danielle Citron accurately observes: "Deepfake technology is being weaponised against women by inserting their faces into porn. It is terrifying, embarrassing, demeaning, and silencing. Deepfake sex videos tell individuals their bodies are not their own – and can make it difficult to stay online, get or keep a job, and feel safe" (Ajder, H. et al., 2019).

The targeted sexualised depiction of women's – and increasingly queer – bodies is not a technical malfunction; it is an expression of patriarchal structures and is being deliberately used. This often occurs in the context

of anti-feminist campaigns and incel movements, which aim to intimidate and exclude certain groups of people (Sittig, J., 2024).

Studies show that over 95% of all deepfakes are sexual in nature. Almost 100% of these depict women. Marginalised groups are particularly affected: queer individuals, Black women and trans women (Paris, B. et al., (2019). This deliberate form of digital violence furthermore creates what is known as a silencing effect (NdM-Glossar, 2025): it distorts digital visibility and restricts democratic participation.

DIGITAL VIOLENCE AS A BUSINESS MODEL

What many still view as a niche phenomenon has become part of a lucrative market. Platforms offering deepfake services typically operate anonymously or from abroad. Access is easy: an email address suffices, payment is made by credit card, Google Pay or cryptocurrency (Meineck, S., 2024). A simple disclaimer ("no editing without consent") is often the only nod to legality. Responsibility is shifted to users, while providers distance themselves from accountability.

However, this very business model could offer a leverage point: The case of Pornhub demonstrates what economic pressure can achieve. In 2020, Visa and Mastercard cut ties with the platform over non-consensual content, prompting significant changes in its upload policies and age verification processes (Der Standard, 2022). A similar mechanism could be applied to deepfake platforms – such as mandatory withdrawal of support from

payment providers, hosting services or search engines that enable their digital infrastructure (Kira, B., 2024).

LEGAL GAPS AND POLITICAL MOMENTUM

But economic pressure alone is not enough. Criminal law has so far struggled to keep pace with deepfake-related offences. Although German law includes provisions such as StGB, 1871, § 201a (violation of personal privacy) and the right to one's own image (Bittner, C., 2021), many deepfake pornography cases fall through the cracks. Legal developments lag behind technological ones, perpetrators remain anonymous, and platforms operate outside EU jurisdiction. The German Women Lawyers' Association (djb) has criticised these legal gaps and called for a dedicated, discrimination-sensitive criminal offence for the unauthorised creation and dissemination of sexualised deepfakes – independent of traditional pornography legislation (Deutscher Juristinnenbund, 2023). There is also an urgent need for targeted training for police and the judiciary, as well as the establishment of a network of specialised prosecutors to raise awareness and develop effective solutions.

EU REGULATION: A FIRST STEP, BUT NOT A CURE-ALL

While national legislation lags behind, developments at the EU level are gaining momentum. The Digital Services Act (DSA) requires platforms to swiftly remove reported illegal content (DSA, Art. 16, 2022) – including deepfakes – where clearly unlawful. The European AI Act introduces transparency obligations (AI Act, Art. 50, 2024) that synthetically generated

content must be labelled as such. And with the new EU directive to combat violence against women (European Parliament and Council, 2024), the non-consensual dissemination of sexualised deepfakes will, for the first time, be criminalised across Europe. Member States – including Germany – have until 2027 to incorporate these rules into national law. In parallel, Germany's proposed Violence Support Act aims to improve access to legal advice and assistance for victims (Die Bundesregierung, 2025).

DIGITAL SELF-DEFENCE AND SOCIAL RESPONSIBILITY

In addition to legal regulation, technical and societal prevention is essential. Tools like Glaze and Nightshade can, for example, alter images in such a way that they become unusable for AI systems – preventing original photos from being repurposed for training datasets or the generation of realistic deepfakes. Think of them as a digital cloak of invisibility against deepfake abuse.

At the same time, public awareness must shift. Image-based sexual violence is still trivialised. Victims are subjected to victim blaming (Friedrich Ebert Stiftung, 2025) rather than support. Yet this is not just about individual fates – it is about structural inequalities that are reproduced and exacerbated in the digital realm.

A COMPLEX PROBLEM DEMANDS MULTIFACETED SOLUTIONS

Sexualised deepfakes are more than technical manipulation. They reflect a shift in digital power dynamics – where existing inequalities are not only reproduced but intensified. The deliberate violation of intimacy and control disproportionately affects those who are already structurally disadvantaged. Deepfakes affect us all – but not equally. That's why we need collective responses that are not merely technical, but feminist, human rights-based and rooted in solidarity. Digital violence is not a fringe issue of internet culture. It is its litmus test.

Organisations like HateAid, the bff – Frauen gegen Gewalt e.V. and anna nackt are already taking action against non-consensual sexualised deepfakes. They support victims, offer contact points, and in 2023 jointlysubmitted a petition to German Digital Minister Volker Wissing, calling for stronger protections and clearer legal frameworks (HateAid, 2023).

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

Bittner, C. (2021): (Fast) keine Fotos ohne Datenschutz. *Stiftung Datenschutz*. https://stiftungdatenschutz.org/ehrenamt/praxisratgeber/praxisratgeber-detailseite/fotos-und-datenschutz-275

Bundesministerium der Justiz (Hrsg.) (1871). *Strafgesetzbuch (StGB)*. https://www.gesetze-im-internet.de/stgb/

Cole, S. (2019): This Horrifying App Undresses a Photo of Any Woman With a Single Click. *Vice*. https://www.vice.com/en/article/deepnude-app-creates-fake-nudes-of-any-woman/

Deutscher Juristinnenbund (2023): Bekämpfung bildbasierter sexualisierter Gewalt. *Policy Paper*, 17. https://www.djb.de/presse/stellungnahmen/detail/st23-17

Der Standard (2024). Skandal um KI-Nacktbilder legt US-Schule lahm. *Der Standard*. https://www.derstandard.de/story/3000000245632/skandal-um-ki-nacktbilder-legt-us-schule-lahm

Der Standard (2022). Nach Vorwürfen: Visa und Mastercard sperren Zahlung von Pornhub-Werbung (2022). *Der Standard*. https://www.derstandard.de/story/2000138125747/nach-vorwuerfen-visa-und-mastercard-sperren-zahlung-von-pornhub-werbung

Die Bundesregierung (2025). Bessere Unterstützung für Gewaltopfer. *Die Bundesregierung*. https://www.bundesregierung.de/breg-de/service/archiv-bundesregierung/gewalthilfegesetz-2321756

European Commission (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on artificial intelligence and amending certain Union legislative acts (AI Act)*, Art. 50. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

European Commission (2022). *Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act)*, Art. 16. https://eur-lex.europa.eu/eli/reg/2022/2065/oj

European Commission (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)*. https://eur-lex.europa.eu/eli/reg/2016/679/oj

European Parliament and Council (2024). Directive (EU) 2024/1385 of 14 May 2024 on combating violence against women and domestic violence. *Official Journal of the European Union*. https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=OJ:L_202401385

Frauen gegen Gewalt e.V. (2025). Hinweise für die Berichterstattung über Gewalt gegen Frauen und Kinder. *Bundesverband Frauenberatungsstellen und Frauennotrufe – Frauen gegen Gewalt (bff)*. https://www.frauen-gegen-gewalt.de/de/ueber-uns/presse/informationen-fuer-die-presse/hinweise-fuer-die-berichterstattung-ueber-gewalt-gegen-frauen-und-kinder.html

Frauen gegen Gewalt e.V. (2024). bff fordert: Bildbasierte Gewalt umfassend bekämpfen. *Bundesverband Frauenberatungsstellen und Frauennotrufe – Frauen gegen Gewalt (bff)*. https://www.frauen-gegen-gewalt.de/de/aktuelles/nachrichten/nachricht/pressemitteilung-bff-fordert-bildbasierte-gewalt-umfassend-bekaempfen-2.html

Friedrich Ebert Stiftung (2025). Victim Blaming. *Friedrich Ebert Stiftung*. https://www.fes.de/wissen/gender-glossar/victim-blaming

Hate Aid (2023): Deepfake-Pornos: Betroffene konfrontieren Wissing. *Hate Aid gGmbH*. https://hateaid.org/wp-content/uploads/2023/10/Pressemitteilung-Schutz-vor-Deepfake-Pornos.pdf

Kira, B. (2024): Deepfakes, the Weaponisation of AI Against Women and Possible Solutions. *Verfassungsblog*. https://doi.org/10.59704/9987d92e2c183c7f

Köver, C. (2023): Gefälschte Nacktbilder von Mädchen sorgen für Aufschrei. *netzpolitik.org*. https://netzpolitik.org/2023/deepfakes-in-spanien-gefaelschte-nacktbilder-von-maedchen-sorgen-fuer-aufschrei/

Maiberg, E. (2025): Chinese AI Video Generators Unleash a Flood of New Nonconsensual Porn. 404 *Media*. https://www.404media.co/chinese-ai-video-generators-unleash-a-flood-of-new-nonconsensual-porn-3/?ref=daily-stories-newsletter

Meineck, S. (2024): Wie Online-Shops mit sexualisierten Deepfakes abkassieren. *Netzpolitik*. https://netzpolitik.org/2024/ki-nacktbilder-wie-online-shops-mit-sexualisierten-deepfakes-abkassieren/

NdM-Glossar (2025): Silencing(-Effekt). *Wörterverzeichnis der Neuen deutschen Medienmacher*innen (NdM)*. https://glossar.neuemedienmacher.de/glossar/silencing-effekt/

Paris, B., Donovan, J. (2019): Deepfakes and Cheap Fakes. The Manipulation of Audio and Visual Evidence. *Data & Society*. https://datasociety.net/wp-content/uploads/2019/09/DS_Deepfakes_Cheap_FakesFinal-1.pdf

Polizei Nordrhein-Westfalen (2021). „Upskirting" und „Downblousing" ist strafbar. *Polizei Nordrhein-Westfalen*. https://polizei.nrw/artikel/upskirting-und-downblousing-ist-strafbar

Reuther, J. (2021): Digital Rape: Women Are Most Likely to Fall Victim to Deepfakes. *The Deepfake Repor*t. https://www.thedeepfake.report/en/09-digital-rape-en

Sittig, J. (2024): Strafrecht und Regulierung von Deepfake-Pornografie. *Bundeszentrale für politische Bildung*. https://www.bpb.de/lernen/bewegtbild-und-politische-bildung/556843/strafrecht-und-regulierung-von-deepfake-pornografie/#footnote-target-25

State of Deepfakes – Realities, Threats, and Impact (2023). *SecurityHero*. https://www.securityhero.io/state-of-deepfak†Aes/#targeted-individuals

Wikipedia (2025). Imageboard. Wikipedia. https://www.fes.de/wissen/gender-glossar/victim-blaming

CAN ŞIMŞEK

AI resistance: Who says no to AI and why?

Artificial intelligence is facing resistance. Around the world, artists, workers and activists are pushing back – not against technology itself, but against the power structures that shape it. These acts of resistance, ranging from data poisoning and creative strikes to protests over data centres, reveal a growing demand for democratic control over AI. In this article, Can Şimşek explains what drives AI resistance and why these acts of pushback could be key to achieving fairer and more accountable AI governance.

READ FULL ARTICLE

PHILIPP MAHLOW & KATHARINA MOSENE

Human expertise for greater fairness in credit lending

Not every credit decision can be left to machines. Banks use automated scoring systems to save time, but real cases are often too complex for algorithms alone. This is where human expertise becomes essential. Philipp Mahlow and Katharina Mosene show how front-desk staff, risk analysts and external agencies work alongside automated models to evaluate creditworthiness. Their analysis highlights why human judgement is vital in ambiguous cases and how opaque AI systems can reinforce bias. How can people and automated processes work together to ensure fairer and more transparent credit decisions?

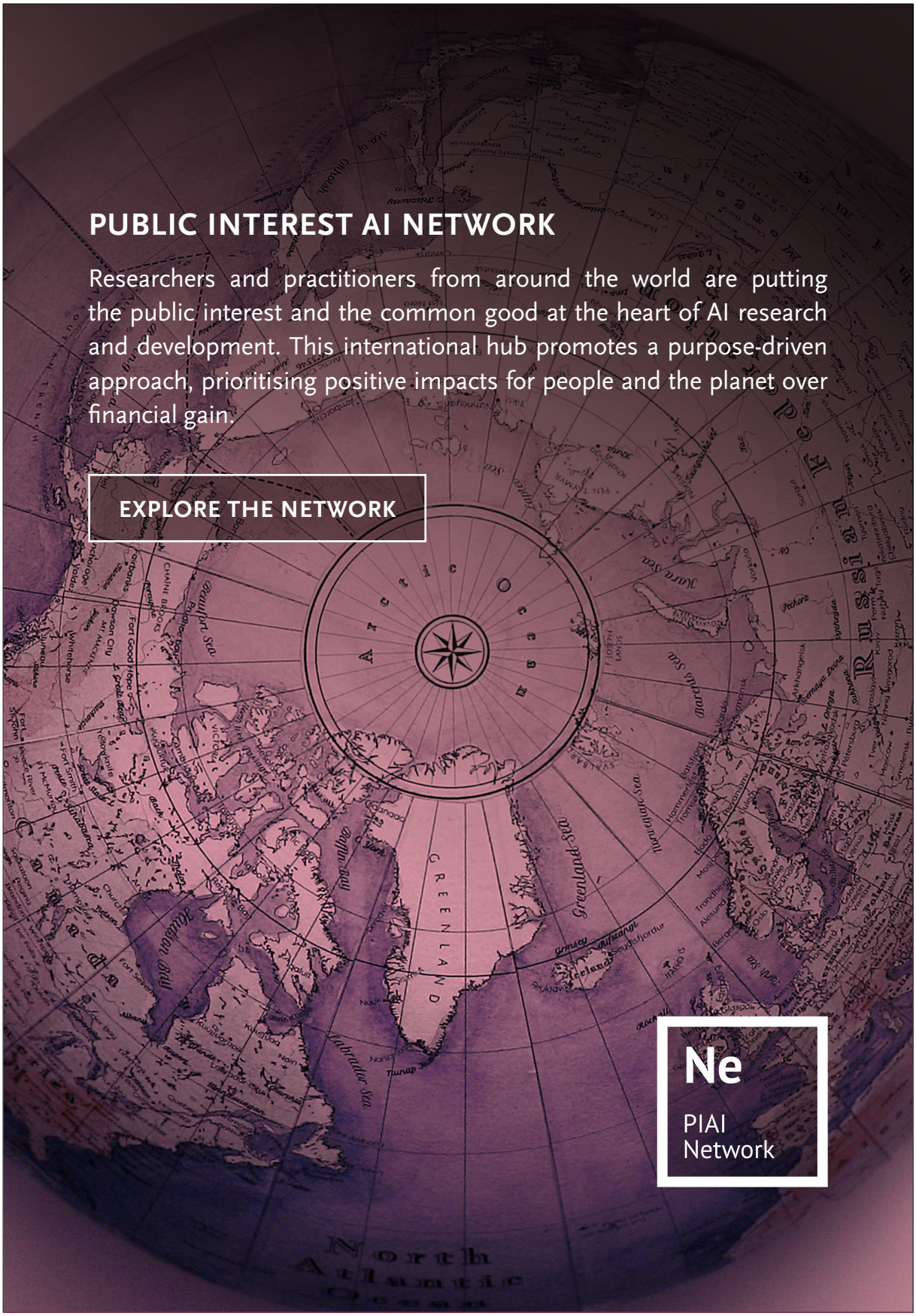
READ FULL ARTICLE

BIRTE LÜBBERT & THERESA ZÜGER

Artificial intelligence with purpose: Mapping the landscape of public interest AI

Does AI always have to be about making a profit? Not necessarily. Innovative projects around the world are using artificial intelligence to address challenges in areas such as climate protection, health, education and democratic participation. Previous research at HIIG mapped over 100 such initiatives in a new dataset. In this article, Birte Lübbert and Theresa Züger demonstrate how this landscape is evolving and identify what is still lacking.

READ FULL ARTICLE



PUBLIC INTEREST AI NETWORK

Researchers and practitioners from around the world are putting the public interest and the common good at the heart of AI research and development. This international hub promotes a purpose-driven approach, prioritising positive impacts for people and the planet over financial gain.

EXPLORE THE NETWORK

Ne
PIAI
Network

AI

DIGITALER SALON

There's no app for that!

Once a month, we publicly discuss the impact of digitalisation on society at Digitaler Salon. In light of the rapid development of new AI systems, the July episode addressed the issue of technologies growing faster than our ability to shape them in line with social values. Is Europe falling behind in the AI race, or should democracy set the pace?

WATCH FULL TALK

ENGLISH SUBTITLES AVAILABLE

ANTHOLOGY

The Realities of Autonomous Weapons

As governments and militaries increase investment in artificial intelligence for defence, questions of accountability, human oversight and the boundaries of machine autonomy become ever more urgent. The way we frame and understand these systems is deeply intertwined with cultural imaginaries, legal norms and political decision-making. This anthology, edited by Thomas Christian Bächle and Jascha Bareis, explores how these realities shape and themselves become shaped by popular culture, regulatory and ethics debates, military doctrines, policies and research.

EXPLORE VOLUME

IMPACT SPOTLIGHT
2025

We foster critical AI literacy in the classroom by giving teachers a ready-to-use method for their lessons in the form of an entertaining card game and workshop.

The impact of research often becomes apparent years after publication. This has certainly been the case with our AI Compass card game. To date, almost 10,000 copies have been requested by schools, universities and institutions across Germany.

Educators have welcomed the game and incorporated it into their activities. For example, the Heinrich-Böll-Stiftung in Lower Saxony has created a special edition to go along with a four-hour workshop concept. These workshops are led by trained student facilitators and offered to schools free of charge. Pupils use the AI Compass as a fun introduction to acquire new knowledge and develop reflective competencies, thereby supporting a better understanding of the societal and sustainability implications of AI.

Due to high demand, the workshop concept is slated to be published next year as an Open Educational Resource, enabling teachers and project teams across Germany to use it in their own classrooms.

Science in times of polarised public issues

Sc

AI

Artificial intelligence and society

Wo

Digital organising and the future of work

FOCUS

Science in times of polarised public issues

The digital society is putting ever more pressure on the creation, sharing and evaluation of scientific knowledge. Universities and research organisations are facing budget cuts and rising public scrutiny, while digital platforms amplify misinformation and accelerate backlash against scholars. These pressures are jeopardising scientific expertise's credibility and ability to inform public discourse, and indeed shaping the conditions under which researchers must engage with society. Sustaining public trust in science is crucial, not only for informed democratic debate – but also for the functioning of democratic decision-making. Our research examines how the challenges to science communication are evolving, particularly in areas of research that become the focus of highly polarised debates. What responsibilities and demands do researchers face when engaging with the general public as well as interest-driven political and economic actors on such contested issues?

VISIT TOPIC OVERVIEW ONLINE

Sc

JEANETTE HOFMANN
INTERVIEWED BY TERESA VÖLKER

Populism, platforms and the challenges of science communication

For the blog journal Elephant in the Lab, Teresa Völker interviewed Jeanette Hofmann about the role of science communication in fragmented publics and the impact of social media on democratic discourse. The conversation also addressed whether researchers can curb the spread of disinformation by engaging differently on social media, or whether these platforms are ultimately unsuitable for meaningful knowledge exchange. So, what constitutes effective communication in today's hybrid media landscape?

What is disinformation and how does the spread of disinformation affect liberal democracies worldwide?

I notice, the longer I work on this topic, that "disinformation" is really an umbrella term – in the sense that it covers many different phenomena, each of which may have its own logic, but they are all grouped under the same label. I find this relevant because I'm particularly interested in a type of disinformation that doesn't appear in most definitions. Take, for example, Michael Hameleer's definition, which describes disinformation as a form of manipulative communication whose underlying intention is often concealed. There's an assumption behind this view that someone is strategically spreading falsehoods in such a way that the audience believes they are receiving accurate information.

This conventional understanding of disinformation is often tied to platforms and algorithms. The assumption is that the attention economy and its algorithms help amplify disinformation. But this definition seems to overlook the kind of disinformation we now frequently see – namely that of political elites. With populist politicians you get the impression they simply don't care about the difference between true and false – at least in the realm of political communication. In other areas, the accuracy of information still matters, say if it will rain tomorrow or if the flight is on time, but politically, accuracy and reliability of information seems to no longer matter. Disinformation is spread with the expectation that people will support certain claims, regardless of whether they believe them or not. It becomes a tool of power: populist politicians act as though they have the authority to define reality.

This kind of disinformation is not hidden, and platforms aren't the primary vectors. It's typically propagated by media outlets like Fox News or other right-wing channels, which bring it to the forefront, only to be taken up and debated by other media. Platforms like X (formerly Twitter) don't necessarily offer the reach needed for such visibility; it is the interaction between fringe media and legacy mass media that amplify these messages to the public.

What has changed in the era of social media?

To be honest, this dynamic has always existed to some extent. What made it explode as a topic was 2016 – with Brexit and the US election. After these unexpected results, many observers sought explanations and blamed the outcome of these votes on platforms, arguing they had spread too much disinformation – and that voters, like "gullible sheep", fell for it and made the wrong choices.

That's how the narrative unfolded. Since then, research and publications on disinformation have increased dramatically. A new research field developed that draws direct connections between disinformation and platforms. That's one level of the discourse. Then there's also a methodological dimension: when it comes to tracing disinformation, platforms are much easier to study than traditional media ecosystems, where drawing connections is much more complex.

If you ask me what empirical role platforms play in spreading disinformation, I'd say they serve as a prelude to populist dynamics. A huge amount of content circulates online, and existing social networks help rumors and

bizarre narratives to spread. So yes, platforms enable a diversity of circulation paths for information. But we can't say disinformation wouldn't exist or would be less visible without platforms. It just would spread differently.

Given that platforms depend on who uses them, do you think scientists and science communicators can help counter the spread of disinformation by engaging differently on social media? Or are these platforms ultimately unsuitable for sharing knowledge?

No, I wouldn't at all say they are unsuitable. Platforms accommodate so many different uses and opinions. That's why many people, especially communication scholars, talk today about a "fragmentation of the public sphere". At the very least, we can observe a kind of differentiation – we now have various coexisting cultural and thematic public spheres. I believe platforms and messaging services are indispensable for science and science communication. Depending on your field, they can be the fastest and most efficient means of communication with colleagues. For many disciplines, they're also an irreplaceable source of information – providing insights that you wouldn't get otherwise. Especially platforms that allow for curating your own feeds and choose which people to follow, create an immense value for its members.

At the same time, we're seeing a withdrawal of many scientific organisations and individual researchers from X (formerly Twitter). There is or used to be the widespread belief that democracy functions best when we all read, hear and discuss roughly the same information. But we also have to

recognise that this kind of homogeneous public sphere only existed for a relatively short time – essentially from the postwar period to the late 1990s. Back then, we had dominant mass media that told relatively unified stories about the world. But if you traveled to another country, you'd immediately notice that entirely different topics and regions dominated the local media coverage. For instance, while Africa might barely appear in German media, it is much more prominent in the UK or France.

This example shows how strong the agenda-setting role of legacy media once was – a degree of selectivity that we no longer experience in the same way today. And that's not necessarily bad; it simply defines a different media era. The real challenge now is: how do we create shared political understanding when everyone sees a different version of reality on their screens? That's the key democratic question in the age of digital platforms.

What does effective communication look like in today's hybrid media environment?

Many people are now working on new participatory formats, recognising that the model of democracy from the past 50 or 60 years may no longer be sufficient. One such idea is citizens' assemblies, where randomly selected participants deliberate on complex policy issues. That's one way to create public discourse – but these formats don't scale easily. And they might not work for every issue, because solving complex problems requires expertise. Random selection alone doesn't always ensure that. What we do see, though, is that when something truly dramatic happens – like the pandemic – we all talk about the same thing again. In moments like that, fragmentation

temporarily reduces, and there's a shared focus. But you can't rely on crises to unify public debate. Perhaps media formats specialised on aggregating information may help us in the future to create a shared public space?

And what role does science play in this context? Is it still the "force of the better argument", or has that faltered?

That was always more of an ideal than a reality. From an academic standpoint, of course it's attractive to see one's own expertise as contributing to better discussions – and ideally to better political decisions. But today, that doesn't seem to really be the case. Quite the opposite: Populist policies seem to have the upper hand. There is more expertise than ever – many people are well-educated and research topics in great detail. Yet when we look at political decisions or statements, they often seem decoupled from empirical evidence. You wonder: "He must know better, or she certainly does – but why say that?" And people who work on expert commissions often leave frustrated. The effort behind gathering and synthesising expert knowledge rarely seems to translate into meaningful policy impact. It's a recurring source of frustration.

Populist politics defines itself as opposition to elites – including knowledge elites. Think of the current health policy in the US, which openly defies established medical standards and remedies. For those aligned with populist thinking, rejecting expert knowledge may feel empowering.

That doesn't mean researchers should withdraw. On the contrary, they must actively defend their role. Research can no longer pretend to be apolitical. Scientists are increasingly drawn into debates, and they need to understand and accept the political consequences of their work.

As a final thought: how should regulation address the use and communication on digital platforms? What challenges and prospects do you see, especially when it comes to protecting freedom of expression?

That's a very complex question. The European Digital Services Act mentions disinformation in passing, but never really defines it. In the actual legal text, the focus is only on illegal content, and it's left up to each member state to determine what qualifies as unlawful. That's been one of the major critiques: member states now have broad power to define content-based offenses, which can be problematic for freedom of speech.

Platforms themselves don't assess disinformation based on meaning. Instead, they use pattern-recognition tools, calling it "coordinated inauthentic behavior". They ask: Who sends which links to whom? How often? Which words pop up? This is because it's very difficult – perhaps impossible – to systematically define and detect disinformation via semantic criteria. What's also often missing from fact-checking debates is the recognition that disinformation isn't always factually false. It's often presented as narratives – blending elements of truth with highly misleading framings or emotionally charged undertones that stir resentment. Fact-checking can only address a small portion of that.

That's why it's so hard to pin it down. I like to refer to the American political theorist Lisa Disch, who argues that disinformation becomes dangerous only when there's no longer a functioning public sphere capable of questioning or correcting it. Her example is the Iraq War – where decisions were based on intelligence reports no one could verify. According to Disch, disinformation is especially dangerous when there's no opportunity for public deliberation about the quality of a statement. And I agree: that's when it becomes a threat to democracy.

My suggestion is: let's clearly define and remove illegal content without weakening freedom of expression – that's straightforward. But beyond that, combating disinformation is more about enabling and safeguarding pluralistic public discourse. A central question here is how to protect quality journalism, especially since platforms dominate the advertising market and the younger generation no longer subscribe to newspapers. We're likely facing the last generation that funds journalism via subscriptions.

So, we need other funding models. In that context, I see the fight against disinformation – not only about takedowns, but as a larger task of rethinking the future of quality reporting, political education and critical media literacy. Studies strongly suggest that good journalism promotes a better quality of public discourse. This also includes science communication.

SAMI NENNO

Misinformation and scientific expertise in talk shows

How would political talk shows change if false claims were corrected immediately? A study by Sami Nenno shows that misinformation isn't limited to social media. German talk shows frequently feature false claims. Over a third of the examined episodes featured misinformation. But when the talk-show guests were scientists, something important happened. They often challenged false claims live, providing crucial counterpoints before the misinformation could take root. Could expert intervention improve the tone and quality of public debate?

READ FULL ARTICLE

ANNA HENSCHEL

When scientists are targeted, what helps?

Is facing hostility towards scientists becoming part of the job? Many researchers now face threats, hate speech and coordinated harassment simply for engaging with the public. Who steps in to help? Anna Henschel reports from a KAPAZ workshop in Berlin where researchers, legal experts and science communicators came together to examine rising hostility towards science. They explored systemic blind spots and discussed what institutions must do to play a more active role in ensuring that researchers stay safe and still be visible.

READ FULL ARTICLE

KATHARINA MOSENE & JOHANNA LEIFELD

Identifying bias, taking responsibility: Critical perspectives on AI and data quality in higher education

Higher education is at a turning point. Artificial intelligence is fundamentally changing how we teach, conduct research and learn. Language models such as ChatGPT and Claude are now part of everyday university life, but their use raises urgent questions. Because AI systems learn from historical data, they can perpetuate existing power structures and inequalities. What does this mean for universities as places of critical thinking and social innovation? The authors address the issue of bias in AI and call for a deliberate, reflective approach to the technology.

READ FULL ARTICLE

FREIA KUPER, BRONWEN DEACON & ANNA AUST

Digital by design, not by default: Resilience in higher education

What does resilience in higher education really look like? And why does it emerge so differently across institutions? Five years after the pandemic, some universities are expanding digital strategies and hybrid models, while others remain committed to analogue-first teaching. By comparing a face-to-face programme in Germany with a digital-by-design Open University in Portugal, the authors of this article examine how resilience can be rooted in both stability and tradition or in flexibility and digital innovation.

READ FULL ARTICLE

RETHINKING SCHOLARLY PUBLISHING

Diamond Open Access means publishing without paywalls for authors and readers. Our new blueprint shows how this non-profit, community-driven model can strengthen fair, transparent and values-based scholarly publishing. This publication presents the findings of the ELADOAH research project and offers practical guidance for libraries, editors, institutions, funders and all those working toward a more equitable open science ecosystem.

EXPLORE BLUEPRINT

SPOTLIGHT ON CURRENT ISSUES

Strengthening democracy against polarisation

What is driving people further apart? The German election in February 2025 has reignited the debate about whether society is becoming increasingly polarised. Many young voters, in particular, have gravitated towards the political right or left. While democracy thrives on diverse opinions, at what point does ideological polarisation become problematic? In March, we explored these pressing questions in a new edition of our HIIG Spotlight, joined by experts Wander Jager and Stefano Braghiroli.

WATCH FULL DISCUSSION

IMPACT SPOTLIGHT
2025

We strengthen the resilience of research organisations by training communication professionals to support researchers and implement strategies for handling hostility towards science.

The KAPAZ project examines the phenomenon of hostility towards science, building the capacity to protect researchers and support research organisations. A key component of this has been the delivery of one-day, practice-oriented train-the-trainer workshops for communication professionals.

Participants were guided in setting up support structures and implementing tried-and-tested strategies within their institutions to respond effectively to hostile or polarised public encounters. They are now implementing these strategies in their own institutions, translating research-based insights into practical action.

The high demand for further sessions, including requests from organisations such as the German Society for Empirical Cultural Studies and the University of Applied Sciences for Police and Public Administration, demonstrates the urgent need for these skills.

Digital organising and the future of work

Wo

Sc

Science communication in times of polarised public issues

Su

Digitalisation and sustainability

FOCUS

Digital organising and the future of work

The dynamic interplay of digitalisation and artificial intelligence (AI) is about to fundamentally reshape the workplace landscape. While there is a consensus that this shift is happening, the precise contours of the impending transformation remain unclear. How exactly will digital organising, digitalisation and AI in the workplace change our day-to-day professional life? In our research, we explore and explain these phenomena, assessing their impacts as well as the associated opportunities and risks for the future of work.

VISIT TOPIC OVERVIEW ONLINE

SONJA KÖHNE

Who hired this bot? On the ambivalence of using generative AI in recruiting

Generative artificial intelligence is advancing into ever more areas of organisational life, and HR practices – particularly hiring processes – are no exception. While AI recruiting technology has long been tied to hopes of freeing up resources for relational work, some evidence points to the contrary. As both recruiters and candidates lean into AI assistance, something essential risks being left behind: the human connection that recruiting depends on. This article is based on research conducted as part of our project: Generative AI in the workplace. It explores how GenAI is reshaping recruitment and why we must pause to ask not just what we can optimise, but why we are optimising in the first place.

"When I prepared this guide for my […] interview […], I ended up with […] a great competency-based interview guide, with behavioural anchors and all sorts of things. And I almost forgot to ask [the candidate]: 'What is actually important to you when looking for a new employer?' […] because the AI didn't generate these questions for me."

—— from an interview with a recruiting manager, 2025

In many areas in and around the workplace, people are currently experimenting with ways of integrating generative AI tools like ChatGPT into their routines and workflows (Dell'Acqua et al., 2023; Retkowsky et al., 2024). The goal is often quite simple: to improve personal effectiveness and save time. For example, few job seekers enjoy writing lengthy cover letters for dozens of applications. In a similar way, few recruiters enjoy scanning dozens of those applications. In comes generative AI, promising both parties support throughout the process. This can lead to significant shifts to a procedure meant to help job seekers and employers evaluate whether or not they will be a good fit for each other – a meaningful task for both (Hunkenschroer & Kriebitz, 2022). It also raises new questions: Do recruiters or hiring managers want to read generic AI-written cover letters? Do job seekers want their applications to be screened by AI? The answer to both may be no. Yet, as generative AI further advances into hiring practices, we are left to ask: What exactly are we optimising for? And at what cost?

THE RISE OF GENERATIVE AI IN RECRUITING

To understand these shifts, it helps to distinguish between the types of AI typically used in people management. Although the line is blurry, scholars often differentiate between discriminative and generative AI. Discriminative AI systems make predictions and classifications, while generative AI systems produce seemingly new content (Feuerriegel et al., 2023; Jebara, 2004). In the context of people management, discriminative AI helps organisations make better personnel decisions (e.g. by predicting candidate-job fit), and generative AI can help to create more effective HR-related content (e.g. images or texts for job ads) (Andrieux et al., 2024). Generative AI qualitatively differs from discriminative AI because it can, among other things, be applied to a broad array of tasks. Thanks to tools like ChatGPT, it is also easily accessible to many (Krakowski, 2025). In people management in general and recruiting in particular, this opens up a wide range of applications (Budhwar et al., 2023; Chowdhury et al., 2024).

THE ALLURE OF AUTOMATION

The introduction of new technology comes with high hopes and (sometimes broken) promises (Garvey, 2018). In the workplace, such promises are often related to automating repetitive tasks to free up resources for more meaningful work. In the case of people management, hopes are often about reducing administrative tasks to free up time for relational or strategic work. For example, if recruiters can use AI to screen resumes more efficiently, they can spend more time on personal interactions with candidates. In interviews

I conducted as part of our project on Generative AI in the Workplace, human resources professionals mentioned further uses specifically of generative AI: to develop interview questions and tasks for work samples, to better tailor job ads to desired target groups, to identify SEO keywords or generate images for job ads, and to write rejection letters. The hope of reclaiming time for more personal interactions with candidates and employees is a consistent theme among human resources professionals. Yet in practice, these hopes often outpace reality, as these tools can disrupt human interactions in subtle but significant ways.

HIDDEN COSTS OF EFFICIENCY

As the quote from the recruiting manager in the introduction suggests, using generative AI in hiring comes with notable risks. While the AI system did help them develop useful interview questions, things can also be lost along the way: They almost forgot to ask the candidate what their expectations were for a new employer! This question can be important to make candidates feel seen and for exploring whether mutual expectations align. Thus, while promising efficiency, generative AI can also diminish aspects of the process that matter deeply or render communication superficial. More broadly, frequent use of AI tools has been linked to declines in critical thinking (Gerlich, 2025). As Nyberg and colleagues (2025) note, verifying simple outputs is relatively easy (e.g. prompting ChatGPT to draft a rejection letter), but these are often the very tasks that were already typically automated before the introduction of generative AI (e.g. through templates or form letters). Verifying outputs requires domain knowledge, yet generative AI

threatens the quality of knowledge in organisations (Retkowsky et al., 2024). Scholars warn that the very use of generative AI for people-related tasks may signal a lack of care for employees, potentially eroding perceptions of interactional justice (i.e. the sense that one has been treated with dignity and respect) (Narayanan et al., 2024; Nyberg et al., 2025). And even when time is saved, our interviews suggest that it remains unclear how that time is actually used.

AN ARMS RACE OF USING GENERATIVE AI?

Job seekers are also turning to generative AI in the hiring process, sometimes in ways that complicate the evaluation of their fit with the organisation and their true expertise. They can now use widely accessible tools like ChatGPT to produce polished resumes and cover letters, prepare ideal responses to likely interview questions, or even use AI teleprompters during virtual interviews that suggest ideal responses in real time (Kwok, 2025). This makes it much harder for recruiters and hiring managers to assess these candidates. As a result, some companies no longer expect cover letters, or they emphasise the importance of in-person interviews. There have also been reports of job seekers sending "AI note takers" to information sessions hosted by potential employers (Ellis, 2024) or using AI to auto-apply to hundreds of job openings at once (Demopoulos, 2024). The outcome can be a "bot versus bot war" where job seekers use AI to send out hundreds of applications, while employers use AI to filter the thousands of similar sounding applications that they receive (Ellis, 2024). Worst case, those screening bots can even show preference for AI-generated applications. The

growing use of generative AI on both sides can feel increasingly absurd and raises the question of whether we are trying to out-automate each other at the cost of authenticity.

RETHINKING THE "HUMAN" IN HUMAN RESOURCES

So, is generative AI a good bot or a bad bot for hiring practices? As with most new technologies, the answer is: It depends. Not on the tool itself, but on how we choose to use it. Generative AI's impact depends on the kind of people management we aspire to build and whether we can align AI with that vision. While tools like ChatGPT can enhance efficiency, particularly in the early stages of hiring (e.g. to optimise job ads), they also risk alienating job seekers through impersonal or literally robotic interactions (e.g. during interviews).

Which brings us back to the question: Who hired this bot? In a sense, we all did – organisations, recruiters and even candidates – often in pursuit of speed, convenience and competitiveness. But in doing so, we may have overlooked the cost of delegating deeply human tasks to machines. The real challenge is not whether to use generative AI or not, but how to use it with intention and care. As HR leaders remind us, the guiding question should not just be what *could* be done with generative AI, but what *should* be done with it (Nyberg et al., 2025). Only then can we ensure that we are not just optimising for efficiency, but for the kind of environments we actually want to work in.

REFERENCES

Andrieux, P., Johnson, R. D., Sarabadani, J., & Van Slyke, C. (2024). Ethical considerations of generative AI-enabled human resource management. *Organizational Dynamics*, *53*(1), 1–9. https://doi.org/10.1016/j.orgdyn.2024.101032

Budhwar, P. et al. (2023). Human resource management in the age of generative artificial intelligence: Perspectives and research directions on ChatGPT. *Human Resource Management Journal*, *33*, 606–659. https://doi.org/10.1111/1748-8583.12524

Chowdhury, S., Budhwar, P., & Wood, G. (2024). Generative Artificial Intelligence in Business: Towards a Strategic Human Resource Management Framework. *British Journal of Management*, *35*(4), 1680–1691. https://doi.org/10.1111/1467-8551.12824

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2023). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. *Harvard Business School Technology & Operations Mgt. Unit Working Paper,* 24 (13), 1–58. https://www.ssrn.com/abstract=4573321

Demopoulos, A. (2024). The job applicants shut out by AI: 'The interviewer sounded like Siri'. *The Guardian*. https://www.theguardian.com/technology/2024/mar/06/ai-interviews-job-applications

Ellis, L. (2024). 'You're Fighting AI With AI': Bots Are Breaking the Hiring Process. *The Wall Street Journal*. https://www.wsj.com/lifestyle/careers/ai-job-application-685f29f7

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative AI. *Business & Information Systems Engineering*, *66*(1), 111–126. https://doi.org/10.1007/s12599-023-00834-7

Garvey, C. (2018). Broken promises and empty threats: The evolution of AI in the USA, 1956–1996. *Technology's Stories*, *6*(1). https://doi.org/10.15763/jou.ts.2018.03.16.02

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, *15*(1), 1–28. https://doi.org/10.3390/soc15010006

Hunkenschroer, A. L., & Kriebitz, A. (2022). Is AI recruiting (un)ethical? A human rights perspective on the use of AI for hiring. *AI and Ethics*, *3*, 199–213. https://doi.org/10.1007/s43681-022-00166-4

Jebara, T. (2004). Generative Versus Discriminative Learning. In T. Jebara,*Machine Learning* (S. 17–60). Springer US. https://doi.org/10.1007/978-1-4419-9011-2_2

Krakowski, S. (2025). Human-AI agency in the age of generative AI. *Information and Organization*, *35*(1), 1–25. https://doi.org/10.1016/j.infoandorg.2025.100560

Kwok, N. (2025). When Candidates Use Generative AI for the Interview. *MIT Sloan Management Review*. https://sloanreview.mit.edu/article/when-candidates-use-generative-ai-for-the-interview

Narayanan, D., Nagpal, M., McGuire, J., Schweitzer, S., & De Cremer, D. (2024). Fairness perceptions of artificial intelligence: A review and path forward. *International Journal of Human–Computer Interaction*, *40*(1), 4–23. https://doi.org/10.1080/10447318.2023.2210890

Nyberg, A. J., Schleicher, D. J., Bell, B. S., Boon, C., Cappelli, P., Collings, D. G., Molle, J. E. D., Feuerriegel, S., & Gerhart, B. (2025). A Brave New World of Human Resources Research: Navigating Perils and Identifying Grand Challenges of the GenAI Revolution. *Journal of Management, 00(00),* p. 1–42. https://doi.org/10.1177/01492063251325188

Retkowsky, J., Hafermalz, E., & Huysman, M. (2024). Managing a ChatGPT-empowered workforce: Understanding its affordances and side effects. *Business Horizons*, *67*(5), 511–523. https://doi.org/10.1016/j.bushor.2024.04.00

GEORG VON RICHTHOFEN

Polished yet impersonal: The unintended consequences of writing your emails with AI

AI-based chatbots are becoming a common tool for writing emails at work. They help professionals write faster, clearer and more politely, saving time and energy for more essential tasks. Yet, this growing reliance raises new questions. In this article, Georg von Richthofen explores the unintended consequences of AI-mediated communication. As everyone from CEOs to interns turns to AI for help with their inbox, what happens to the way we think, connect and express ourselves? Are we outsourcing more than just words?

READ FULL ARTICLE

PHILIP MEIER

AI at the microphone: The voice of the future?

The audio industry is in flux. Podcasts, audiobooks and audio dramas are more popular than ever; meanwhile, artificial intelligence is transforming the production and distribution of this digital audio content. Voice synthesis tools can now automatically narrate texts and algorithms can generate entire podcast episodes. What new opportunities are emerging, and what challenges lie ahead? In this article, Philip Meier explores the most exciting trends and examines the present and future of AI in the audio world.

READ FULL ARTICLE

LORENZ GRAF-VLACHY

Emotionless competition at work: When trust in artificial intelligence falters

In many workplaces, humans already work side by side with artificial intelligence. Systems such as ChatGPT and DALL–E, as well as analytical tools, support decision-making processes, provide creative input and handle complex tasks – often more quickly and accurately than humans. However, this superiority has its downside. Research by Lorenz Graf-Vlachy shows that those who feel outperformed by AI begin to doubt not only their own abilities, but also the technology itself. Paradoxically, a machine's strength can undermine trust in it. When introducing AI, what should companies keep in mind?

READ FULL ARTICLE

DIGITALER SALON

Rausreguliert?

Start-ups are driving innovation with AI, but the EU's new AI Act is set to change the rules of the game. How can small teams comply with the Act's complex documentation and risk control requirements without compromising their creativity? In the May issue of our Digitaler Salon, we discussed whether the new regulation provides entrepreneurs with legal certainty or hinders innovation.

WATCH FULL TALK

ENGLISH SUBTITLES AVAILABLE

STRENGTHENING DIGITAL SKILLS

Micro-credentials on AI and data

Explore our new online courses! The first introduces you to the technical foundations of AI, encouraging you to think critically about its application in various contexts. The second course explains how data is produced, governed and negotiated. These two micro-credentials have been developed to prepare students from Bangladesh, India, and Vietnam without a technical background for the rapidly evolving demands of the digitalised work environment.

ENROL IN COURSES

IMPACT SPOTLIGHT
2025

We empower the next generation to thrive in the digital workplace by giving them a practical understanding of AI and data, as well as their societal and ethical implications.

The Strengthening Digital Skills through Micro-Credentials project develops flexible courses to equip university students in Bangladesh, India and Vietnam with industry-aligned skills.

The two training units, Decoding AI and Understanding Data, teach participants to grasp key concepts, critically assess technological systems and understand social and regulatory implications, including risks such as automation bias. In collaboration with the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) and partner universities, HIIG designed these micro-credentials to require no prior technical knowledge and to actively support female students' participation in digital fields.

The courses are already being adopted by partner universities in Asia and integrated into regular teaching. This demonstrates how research-based insights can be translated into practical skills that foster inclusive participation in the digital economy.

Digitalisation and sustainability

Su

Wo

Digital organising and the future of work

FOCUS

Digitalisation and sustainability

When it comes to sustainability, digitalisation opens up great opportunities but poses myriad challenges. It is important that societies, organisations and individuals begin to grasp the sustainable potential of digital technologies. Currently, their negative effects often outweigh the positive ones due to enormous requirements for material and energy, but also due to discrimination against certain social groups – the same ones affected most by the negative consequences of outsized material and energy consumption on the whole. In our research at HIIG, we investigate how digital and sustainable transformation can be shaped together. How can technology be introduced and used in a responsible manner – from a social, economic and ecological perspective?

VISIT TOPIC OVERVIEW ONLINE

LENA WINTER & THERESA ZÜGER

Blind spot sustainability: Making AI's environmental impact measurable

Efficient, smart and environmentally friendly? Artificial intelligence is often hailed as the solution to many of the major challenges we face today, including climate change. However, behind this optimistic vision of the future lies a blind spot: AI consumes vast quantities of energy, produces CO_2 emissions, and its environmental footprint remains largely opaque. However, reliable data, suitable measurement methods and binding standards are still lacking. In this article, Lena Winter and Theresa Züger explain what needs to change in order to assess the impact of AI.

Every stage of an AI system's life cycle demands vast resources: from hardware manufacturing and data centre construction to the development and training of AI models and their subsequent use. At the end of this chain lies heaps of outdated hardware e-waste. All of these steps require rare earths, energy and water, and must be included in AI sustainability assessments (Smith & Adams, 2024).

It's important to note that sustainability has many facets – ecological, social and economic. This blog post focuses on environmental protection – in other words, on the ecological dimension. It's about how we can conserve resources and preserve nature.

AI AND ECOLOGICAL SUSTAINABILITY – WHAT WE (DON'T) KNOW

For a long time, AI was seen as a technological beacon for the green transition. Increasingly, however, its substantial environmental impact is being discussed in more critical terms. We highlighted how the public discourse around AI is shifting in Germany already in our Digital Society Blog (Liebig, 2024).

Little is known about AI's actual total resource consumption. Concrete data and figures remain scarce, often kept under wraps. Major AI providers and data centre operators – over a third of which are based in the US (Hajonides et al., 2025) – like Google or Meta, offer little transparency about their actual usage. As a result, AI's ecological footprint can currently only be estimated rather than precisely measured (Smith & Adams, 2024). To make matters worse, we lack standardised methods to reliably assess AI's

environmental impacts across its entire life cycle (Kaack et al., 2022). A comprehensive evaluation must include both resource consumption and resulting emissions, such as those from power generation to run data centres (Smith & Adams, 2024).

AI FOR ENVIRONMENTAL PROTECTION?

A simplistic black-and-white view of AI isn't helpful here. There are projects that deliberately use AI to protect the environment. On our Digital Society Blog, we've presented such examples. For instance, AI can help detect leaks in wastewater systems, thus protecting drinking water and ecosystems from contamination. It can also be used to identify and preserve habitats of endangered species (Kühnlein & Lübbert, 2024).

But these projects also rely on the same resource-intensive technologies. This makes it difficult to gauge their actual environmental benefit. We lack robust data to assess the environmental gains versus the resources consumed over the entire life cycle. So, do the trade-offs render these solutions unsustainable?

MEASURING AI'S ENVIRONMENTAL IMPACT

This is where the new research project Impact AI: Evaluating the impact of AI for sustainability and public interest comes in. The project is run by the Alexander von Humboldt Institute for Internet and Society (HIIG) in collaboration with Greenpeace and the Economy for the Common Good.

Over five years, the project will examine 15 AI initiatives from various sectors. Its goal: to systematically and holistically assess their real impact on society and the environment. A new methodology is being developed that combines indicators such as energy efficiency and AI-generated emissions with a qualitative evaluation of ethical and social dimensions. This approach aims to make both the sustainability *of* AI and sustainability *through* AI visible. It helps identify the potential and strengths as well as the limitations of AI projects that seek to contribute to sustainability and the public interest.

Both in terms of evaluating how sustainable AI systems themselves are, and their contribution to environmental goals, there's still a lack of clear data or criteria. This presents a challenge not only for conscientious end users but particularly for organisations aiming to develop AI in a responsible and sustainable way.

WHAT DOES SUSTAINABLE AI LOOK LIKE?

How can we align AI with ecological sustainability? Initial ideas were developed during a workshop at the conference Yes, we are open?! Designing Responsible Artificial Intelligence, organised by the Berlin University Alliance, the University of Vienna, Wikimedia, the Weizenbaum Institute, and the HIIG. The event focused on the intersection between open knowledge, AI and science. A key question: To what extent does open access to research findings and data influence fair and sustainable AI development?

In a discussion moderated by HIIG, participants from academia, civil society and NGOs jointly formulated policy recommendations aimed at advancing the discourse on AI's ecological responsibility.

A MONITOR FOR GREATER AWARENESS?

One such recommendation focused on AI systems' resource consumption: How can it be made more transparent – especially for users who want to weigh the benefits of AI use against its environmental costs? Would people use ChatGPT or other AI tools as frequently if they knew that a single chatbot conversation can consume up to 500 ml of water (Li et al., 2023)?

This kind of direct feedback – similar to the "flight shame" phenomenon – could encourage a more critical perspective on individual AI use. However, for people who rely on AI in their daily work – to be more productive, generate content faster or automate decisions – there may be little real choice to opt out.

Individualising the problem, however, risks shifting the burden. It foregrounds users' responsibility for AI sustainability, while structural levers, such resource consumption disclosure or environmental protection enforcement, remain in the background.

So, consumption monitoring may not be a silver bullet, but it's a tool to raise awareness about the link between AI use and resource demand. And that awareness is a critical foundation for moving the public debate on AI's ecological consequences forward.

THE ELEPHANT IN THE ROOM: LACK OF TRANSPARENCY, MISSING DATA

Developing an accurate consumption monitor still faces one major hurdle: a lack of reliable data and transparency about AI's environmental effects.

The discussion group quickly reached consensus that independent measurement is needed. A key lever: greater insight into the data centres that run AI systems. How much computing power is used for AI? Where are the servers located? What are the energy sources? How much water is consumed? Most of these questions remain unanswered, simply because operators don't disclose the data – yet.

The Data Centre Registry created under the EU Energy Efficiency Directive holds enormous potential. It aims to establish a European database for data centres. In Germany, operators are now required to register and annually report information on energy use and heat recovery to the Federal Ministry for Economic Affairs and Climate Action. However, it's still unclear how much of that computing power goes specifically to AI.

Thus, calls for comprehensive reporting and documentation standards persist. These must be uniform and holistic to assess and compare environmental impacts across AI's life cycle. Moreover, measurement must not be left to the industry alone to self-monitor. To avoid greenwashing, independent or public entities must oversee those assessments.

AI POLICY IS SUSTAINABILITY POLICY

To implement these demands, a change in mindset is necessary. The risk posed to ecological sustainability by AI must be recognised by policymakers. High-resource AI systems raise questions of responsibility – and that makes AI policy also environmental policy.

What could legislators do? Existing environmental assessment tools and incentive systems could be expanded and applied to AI. This includes comprehensive life-cycle assessments for digital services. Appropriate tools are already available in the construction industry. But to do this, the entire digital supply chain – everything required to provide AI systems – must be disclosed and factored in. Additionally, carbon pricing ought to be extended to digital services, especially those provided outside Europe. That way, emissions from non-European data centres would also be accounted for. While mechanisms for carbon border adjustment exist within the EU, they currently only apply to products like steel or fertilizers.

LOOKING AHEAD

There was also a sense of frustration in the discussion. Many participants criticised the slow pace of political processes and the lack of serious sustainability thinking in AI deployment. One question came up over and over: How can individuals and organisations take responsibility themselves?

Yet, there a feeling of momentum also presided. Participants were motivated to jointly push for more transparency in AI's environmental impact. The

idea is to strengthen public discourse and ensure that AI and environmental policy are increasingly seen as interconnected. Science can play a key role here by developing better methods to evaluate AI's resource use and making them accessible.

From this shared concern, a new network has emerged: "AI and Sustainability". Researchers and civil society representatives have come together to regularly exchange ideas, critically monitor developments and propose concrete actions. Their goal: to place AI's ecological responsibility permanently on the political and societal agenda – not someday, but now.

REFERENCES

Hajonides, J., McCarthy, J., Koulouri, K., & Camargo, R. (2025) Navigating AI´s Thirst in a Water-Scarce World. A Governance Agenda for AI and the Environment. *NatureFinance*. https://www.naturefinance.net/resources-tools/navigating-ais-thirst-in-a-water-scarce-world

Kaack, L. H., Donti, P. L., Strubell, E., Kamiya, G., Creutzig, F., & Rolnick, D. (2022). Aligning artificial intelligence with climate change mitigation. *Nature Climate Change*, 12(6), Article 6. https://doi.org/10.1038/s41558-022-01377-7.

Kühnlein, I., & Lübbert B. (2024). Ein kleiner Teil von vielen – KI für den Umweltschutz. *Digital Society Blog*. 10.5281/zenodo.13221001.

Li, P., Yang, J., Islam, M. A., & Ren, S. (2023). Making AI Less „Thirsty“: Uncovering and Addressing the Secret Water Footprint of AI Models. *arXiv*. http://arxiv.org/abs/2304.03271.

Liebig, L. (2024). Zwischen Vision und Realität: Diskurse über nachhaltige KI in Deutschland. *Digital Society Blog*. doi: 10.5281/zenodo.14044890.

Smith, H., & Adams, C. (2024). Thinking about using AI?. Here's what you can and (probably) can't change about its environmental impact. *Greenweb Foundation*. Retrieved 14.04.2024, from https://www.thegreenwebfoundation.org/publications/report-ai-environmental-impact

NELE BUSS

The digital metabolic rift: Why do we live beyond our means online?

Our digital habits leave a bigger carbon footprint than we think. Although clicks, streams and AI prompts may feel weightless, they consume real energy, water and resources. Invisible infrastructures make the environmental cost of digital consumption easy to ignore. So, while we recycle and avoid flying to protect the planet, our ecological awareness often ends where the digital begins. Nele Buß introduces the concept of a "digital metabolic rift" and explains how it can help us rethink the environmental impact of our digital activities.

READ FULL ARTICLE

ANNE-KATHRIN BERG & FREYA HEWETT

Online echoes: the Tagesschau in Einfacher Sprache

Simplified news isn't simple at all. Since June 2024, the German Tagesschau has offered a simplified language edition of its news broadcast to improve its own accessibility. Right after its launch, the so-called Tagesschau in Einfacher Sprache sparked a heated online debate. Anne-Kathrin Berg and Freya Hewett analysed Reddit discussions to identify six recurring themes that shed light on how we negotiate participation in the public sphere. What can public broadcasters learn from this online echo?

READ FULL ARTICLE

ANNIKA ULICH & THOMAS SCHILDHAUER

Opportunities to combat loneliness: How care facilities are connecting neighborhoods

Our society is getting older and loneliness is one of the biggest challenges of demographic change. As mobility declines and social spaces disappear, care facilities are reconsidering their role. Annika Ulich and Thomas Schildhauer demonstrate how they are evolving into open neighbourhood hubs that promote social and digital participation. These new hybrid meeting places offer digital skills training and community-focused services, aiming to strengthen community ties and forge meaningful connections across generations.

READ FULL ARTICLE

PAUL VILCHEZ

Impactful by design: A call for digital entrepreneurs driven to create positive societal impact

Good intentions can still build bad technology. In this article, Paul Vilchez explains why digital solutions often fail when they ignore the lived realities of the local communities they target. Looking at a healthcare app developed for Ethiopia as an example, he demonstrates how participatory and value-sensitive design can create impactful tools that truly serve people. How can entrepreneurs embed societal values from the very first prototype?

READ FULL ARTICLE

MAURICE STENZEL

Escaping the digitalisation backlog: data governance puts cities and municipalities in the digital fast lane

German cities and municipalities perform many functions that facilitate our daily lives: they organise traffic, manage schools and childcare centres, and provide social services. Many of these tasks could be improved and simplified by using digital solutions, such as user-friendly transport apps, digital public office services and data-based traffic planning to reduce congestion. However, for such digital solutions to realise their potential, data from different sources must be meaningfully brought together. In practice, this often entails considerable challenges. Projects that use data often fail in the planning phase and become part of the digitalisation backlog. We have investigated the reasons for this and developed a digital handbook based on our research: the Data Governance Guide. This handbook supports German administrations in implementing data-driven projects in a structured and legally compliant manner, ensuring that digital solutions are adopted and provide tangible benefits for citizens.

WHAT IS THE PURPOSE OF DATA GOVERNANCE?

Imagine if local public transport adapted its timetable according to the number of passengers each day. It could then deploy larger vehicles or increase the frequency of services if demand increased. This is technically feasible, and the necessary data is available in principle. For a project like this to be implemented, however, many parties would need to work together. Transport companies would have to provide data on routes, vehicles and passenger flows, while the city would need to gather and forward data on the traffic situation. A technology provider would also need to develop the planning software. This cooperation must be designed so that all laws, particularly those relating to data protection, are complied with. This requires clear regulations: Who is allowed to process which data? How may it be used? How will it be protected, and how long may it be stored for? What requirements must the data meet so that administrations can use it to make decisions?

This is where data governance comes in. It establishes rules for data processing, allocates responsibilities and organises the process. It also ensures that legal certainty is always guaranteed and that everyone involved ultimately benefits. Digital innovations should not just be technically possible, but indeed should also improve our lives. Although the goal and the rules are often clear, many digital projects are not getting off the ground. They are stuck in a digitalisation backlog. Why is that the case?

WHY IS IMPLEMENTING DIGITAL SOLUTIONS IN CITIES AND MUNICIPALITIES OFTEN SO DIFFICULT?

There are many reasons why digitisation projects grind to a halt. The most important reason is that administrations, companies and citizens interpret the value of data differently depending on the use case. For some, it is an opportunity to do something useful; for others, it poses risks. For data to be used, the benefits must outweigh the risks.

The problem with this is that the risks of data protection or the disclosure of business secrets are often immediately apparent when sharing data with third parties. The benefits, on the other hand, often only become apparent later – for example, when a flexible bus timetable has reliably increased passenger numbers. Another problem is that the required data is distributed across many different bodies, including administrations, companies, research institutions and, increasingly, the general public. As benefits and risks assessments change from case to case, and as laws and technology are constantly being adapted, flexible solutions are needed. Many existing data governance concepts are overly complicated and technical, and therefore do not work well in practice.

For successful implementation, the parties involved need to harmonise three areas. Firstly, they must coordinate their processes when working together on a specific project. Secondly, they must ensure that digital solutions function smoothly. Thirdly, they must ensure that legal requirements are complied with. So far, there has been a lack of approaches that allow local employees to implement this independently. Solutions need to go beyond

one-off measures and be made available more broadly in more municipal administrations and communities, and, crucially, be uncomplicated.

HOW THE DATA GOVERNANCE GUIDE CAN HELP

We developed the Data Governance Guide as part of our Data & Smart City Governance project. This digital handbook is aimed directly at administrations, companies, politicians and citizens. It provides cities and municipalities with practical, three-step instructions for planning and implementing data projects in a comprehensible manner that takes legal requirements into account. The guide was developed with a practical focus and tested in various areas, including air quality management in Berlin and the digital strategy in Haßfurt.

"Measures must be planned according to need. But when is a measure appropriate to the need? This is where the guide comes in. It enables local authorities to track processes and planned measures in a structured manner, record them clearly and highlight processes, actors, data and associated potential at key points."

– Thorsten Kempf, Chief Digital Officer model project Smart Green City Haßfurt

The aim was to investigate how public administrations, companies and civil society could collaborate to use and process data to help achieve cleaner

air. Such projects have enabled us to demonstrate that this approach is not only theoretical but also practical.

The guide shows how data-driven projects can be better positioned right from the start in three simple steps.

Understanding processes: first of all, it helps to make processes visible and clarify responsibilities. Our research has shown that administrative processes are often not documented transparently. This results in unclear responsibilities and legal uncertainties when handling data, which slow down many projects at an early stage.

Exploiting data: the second step is to determine what data can be used at which points in the process. Only then does one look for who has this data, how it can be obtained and which legal, organisational or technical issues need to be considered.

Initiate change: the third step involves engaging stakeholders in a targeted manner to build trust, resolve conflicts early on, and communicate the value to all parties. Many projects come to an early end because it is often difficult to convince administrative employees of the concrete benefits of digital projects, and it becomes even more difficult when external stakeholders such as companies, research institutions or citizens are involved. Although there are formal participation procedures, these often do not provide sufficient opportunity to understand the different expectations regarding the benefits and risks, or to develop joint solutions. Consequently, conflicts usually only become apparent at a late stage and are then difficult to resolve.

The guide provides administrative staff, companies, politicians and citizens with a variety of templates, checklists and concrete examples to successfully navigate these three phases. It provides straightforward answers to legal, organisational and technical questions such as: How does the process work, and who is involved? What data is involved and how is it processed? How can we ensure that everything complies with legal requirements and that the benefits outweigh the risks? Who can get involved in the process and help decide when and how? It turns abstract theory into a practical tool that supports local authorities in implementing their digital ideas.

HOW DATA GOVERNANCE CAN PUT CITIES AND MUNICIPALITIES ON THE FAST TRACK TO THE DIGITAL AGE

Digital projects can be successfully implemented if everyone involved works together in a well-coordinated manner and the right structures are in place. Data governance ensures that digitalisation is not hindered by uncertainties and ambiguities, but can instead realise its potential. This means that politicians and administrators receive a better basis for decision-making, while citizens save time thanks to digital solutions and have easier access to administrative services. Companies can utilise their digital solutions in cities because they understand the requirements of municipalities when working with data. The Data Governance Guide supports this, showing how diverse interests can be transformed into a shared approach that engages and involves all stakeholders. This makes digital change a tangible development that improves life in cities and municipalities, rather than remaining an abstract vision.

DIGITAL HANDBOOK

Data Governance Guide

German administrations hold valuable data that could improve mobility, public services and sustainable urban development – yet much of its potential remains unused. This digital handbook offers a practical introduction for local authorities to plan and implement their own data-driven projects. Its goal: helping cities and municipalities overcome the digitalisation backlog.

EXPLORE THE GUIDE

DIGITALER SALON

Watt für'n Prompt

Behind every chatbot request are large data centres that consume massive amounts of energy and water. In this issue of Digitaler Salon, experts discussed the true scale of AI's demand for resources, the global inequalities this creates, and the shared responsibility of companies, policymakers and users. Can AI ever be climate-friendly?

WATCH FULL TALK

ENGLISH SUBTITLES AVAILABLE

IMPACT SPOTLIGHT
2025

We make Germany's cities smarter and future-ready through effective data governance in administrations.

The Data & Smart City Governance project has investigated why data-driven projects in German municipalities often fail to come about. Legal uncertainties and unclear responsibilities, structures and processes prevent digital innovations from moving beyond the planning phase.

To change this, we developed the Data Governance Guide. It helps the various actors make sustainable use of the enormous potential of data and integrate it into their processes in a legally compliant manner.

Tested from the beginning with public administration practitioners and tailored to their needs, the handbook has been received with great enthusiasm by chief digital officers, administrative managers and other municipal multipliers in workshops and training courses. While the full integration of these skills and processes into municipal administration will unfold in the coming years, municipal authorities have already begun applying the guide in their workflows.

EMPOWER RESEARCH TO INSPIRE CHANGE

Help amplify Europe's voice in the global dialogue on the internet and society. By supporting us, you contribute to advancing scientific research and fostering public engagement. Addressing societal challenges starts with solid foundational research – your support ensures we can continue to explore and critically shape the digital transformation. Together, we can make a lasting impact.

SUPPORT OUR WORK

Alexander von Humboldt
INSTITUTE FOR INTERNET
AND SOCIETY

encore

Volume 2025/2026
Published December 2025

PUBLISHER

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

EDITORIAL TEAM

Frederik Efferenn, Freia Kuper,
Katharina Mosene, Björn Scheuermann,
Larissa Wunderlich, Sarah Ziedler

DESIGN

Larissa Wunderlich

DOI 10.5281/zenodo.17912843