

# “AI will fix this” – The Technical, Discursive, and Political Turn to AI in Governing Communication

Big Data & Society  
July–December: 1–8  
© The Author(s) 2021  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/20539517211046182  
journals.sagepub.com/home/bds  
 SAGE

Christian Katzenbach<sup>1,2</sup>

## Abstract

Technologies of “artificial intelligence” (AI) and machine learning (ML) are increasingly presented as solutions to key problems of our societies. Companies are developing, investing in, and deploying machine learning applications at scale in order to filter and organize content, mediate transactions, and make sense of massive sets of data. At the same time, social and legal expectations are ambiguous, and the technical challenges are substantial.

This is the introductory article to a special theme that addresses this turn to AI as a technical, discursive and political phenomena. The opening article contextualizes this theme by unfolding this multi-layered nature of the turn to AI. It argues that, whereas public and economic discourses position the widespread deployment of AI and automation in the governance of digital communication as a technical turn with a narrative of revolutionary breakthrough-moments and of technological progress, this development is at least similarly dependent on a parallel discursive and political turn to AI. The article positions the current turn to AI in the longstanding motif of the “technological fix” in the relationship between technology and society, and identifies a discursive turn to responsibility in platform governance as a key driver for AI and automation. In addition, a political turn to more demanding liability rules for platforms further incentivizes platforms to automatically screen their content for possibly infringing or violating content, and position AI as a solution to complex social problems.

This article is a part of special theme on The Turn to AI. To see a full list of all articles in this special theme, please click here: <https://journals.sagepub.com/page/bds/collections/theturntoai>

## Introduction

When Facebook CEO Marc Zuckerberg was pressed in the 2018 Senate Hearing upon issues of misinformation, hate speech and privacy, he was eager to present a solution: “AI will fix this!” (Katzenbach, 2019). In this hearing, senators asked Zuckerberg not only about what had happened in previous years, but also demanded to hear about the company’s plans to respond adequately and responsibly in the future to the challenges posed by disinformation campaigns, the spread of hate speech, terrorist propaganda and other problematic content. In different phrases, Zuckerberg repeatedly referred to the development and increasing use of AI-powered systems to detect hate speech, terrorism and misinformation: “In the future, we’re going to have tools that are going to be able to identify more types of bad content.”<sup>1</sup> He reassured the senators that future systems will also cope much better with the difficult contextual and nuanced classification of language:

“Over a 5 to 10-year period, we will have A.I. tools that can get into some of the nuances – the linguistic nuances of different types of content to be more accurate in flagging things for our systems. But, today, we’re just not there on that.” Whatever the challenge, Zuckerberg positioned new technology as the answer to complex social challenges.

Zuckerberg is not alone with this resort to technology. Technologies of “artificial intelligence” (AI) and machine learning (ML) are regularly presented as solutions to key

<sup>1</sup>Centre for Media, Communication and Information Research (ZeMKI), University of Bremen

<sup>2</sup>Alexander von Humboldt Institute for Internet and Society (HIIG), Berlin

### Corresponding author:

Christian Katzenbach, Centre for Media, Communication and Information Research (ZeMKI), University of Bremen, Linzer Str. 4-6, 28359 Bremen, Germany  
Email: [katzenbach@uni-bremen.de](mailto:katzenbach@uni-bremen.de)



problems of our societies. Companies are developing, investing in, and deploying machine learning applications at scale in order to filter and organize content, mediate transactions, and make sense of massive sets of data. At the same time, social and legal expectations are ambiguous, and the technical challenges are substantial. In particular, addressing issues such as misinformation and hate speech with AI technologies evokes particular problems and harms, since the contextual nature of these types of content limits the accuracy of traditional algorithmic systems and thus augments harms such as overblocking. Machine learning technologies might indeed better fit with fuzzy distinctions between legitimate and illegitimate content, yet observation of the rapid deployment of AI technologies in other sectors suggests that problems of equality and diversity, discrimination and bias are often rather amplified by automated technologies than diminished. Thus, it is of utter importance to integrate social and legal expertise into the debate about how technologies reorder communication and society – and how (and if) they should be applied to thorny social problems at scale.

In this context, this special theme for *Big Data & Society* investigates the rapid turn to AI in ordering communication online. The resulting set of research articles and commentaries draws on perspectives from multiple disciplines ranging from media and communications to law and computer science. This introductory article contextualizes this theme by identifying the multi-layered nature of this turn. Whereas public and economic discourses position this development as a technical turn to AI with a narrative of revolutionary breakthrough-moments and of technological progress (Bareis and Katzenbach, 2021), the widespread deployment and normalization of AI and automation in the governance of digital communication is at least similarly dependent on a parallel discursive and political turn to AI. This opening article develops this argument by firstly positioning the current turn to AI in the longstanding motif of the “technological fix” in the relationship between technology and society. The piece will, secondly reconstruct the discursive turn to AI, and, thirdly, identify a political turn to AI in governing communication, before introducing the articles of this special theme.

## The technological Fix – positioning technologies as solutions to social problems

With his reaction at the Senate hearing 2018, Facebook CEO Marc Zuckerberg resorted to a regular motif in the relationship between technology and society, which actors from business and technology powerfully introduce into discourses time and again: the “technological fix”. This motif positions technology as a necessary and functional solution to social problems and challenges. Rudi Volti pointed out the ubiquity

of this motif in the 1990s: “The list of technologies that have been or could be applied to the alleviation of social problems is an extensive one, and examples could be supplied almost indefinitely. What they have in common is that they are ‘technological fixes’, for they seek to use the power of technology in order to solve problems that are nontechnical in nature.” (Volti, 2014, p. 30) Blind faith in the effect of technology is usually combined with ignorance of the causes and dynamics of existing social problems: “In this view, traffic management systems cope with the increasing number of cars in cities (and car traffic is not reduced), food imports keep the poorest from starving (and the causes are not combated), cattle are culled (and industrial factory farming is not adopted)” (Degele, 2002, p. 25, translation by the author).

In the context of digitalization, Zuckerberg is by no means the only one to carry on this narrative – rather, it is a central motif of the basic narrative traits of the US-dominated digital industry (Daub, 2020). Evgeny Morozov, for example, describes how Silicon Valley companies treat social problems or complex contexts such as health and mobility as problems that can be solved functionally. By providing ever new services and apps, they promise to optimize processes and social interactions. What Volti calls “technological fix”, Morozov calls “solutionism”: “Recasting all complex social situations either as neat problems with definite, computable solutions or as transparent and self-evident processes that can be easily optimized – if only the right algorithms are in place!” (Morozov, 2013, p. 5).

The current turn towards addressing complex questions of shaping and ordering public communication through algorithms and artificial intelligence is thus by no means unique, but is part of a long-lasting motif in the relationship between society and technology. This finding is important for the classification of the current algorithmic turn, as it already indicates that this turn cannot be explained by technical progress alone. Rather, it is a recurring motif that gains or loses importance depending on the social problem situation. The impetus for the prominent positioning of a technological fix as a powerful pattern of interpretation and explanation can be triggered by technical impulses, but it can carry nor explain such a development alone. So while there are clearly technological advancements in identifying and classifying content (Gorwa et al., 2020; Cardon et al., 2018), there is more to this turn to AI than technological progress. In the following, we thus reconstruct a discursive, and then a political development that both foreground the role of AI in improving society and governing communication online.

## The discursive turn to Ai – answering a turn to responsibility with the technological fix

On the discursive level, a conjunction of two developments enables the rise of AI and automation as a seemingly

adequate solution to a complex social problem. The first is the massive general increase of attention and visibility for AI in public discourse since the mid-2010s. Existing studies of media reporting identify that new products and supposed innovations clearly dominate the coverage, with business actors and tech entrepreneurs featuring much more often in AI reporting than other stakeholders (Brennen, Howard, and Nielsen, 2018; Chuan, Tsai, and Cho, 2019; Fast and Horvitz, 2017; Puschmann and Fischer 2021). This reporting style builds on longstanding narratives that attribute magical properties to technologies, and specifically to AI (Bory, 2019; Cave and Dihal, 2019). Even governmental strategies and communication adopt this narrative by positioning AI as an inevitable and massively disrupting technological development with high economic opportunities (Bareis and Katzenbach, 2021; Zeng, Chan and Schäfer, 2020). This “enchanted determinism” in the general AI discourse (Campolo and Crawford, 2020) constitutes a particularly strong expression of the technological fix motif.

This general AI discourse functions as a sounding board for the second discursive development relevant here, and that is the remarkable shift in platform governance discourse since 2015 – 16: for a few years now, public and political actors have been increasingly demanding that platforms take responsibility for the content and communication dynamics on their services (responsibility turn). As a consequence, platform operators are responding to this growing pressure with the promise of a technical solution. In this context, the general AI discourse greatly eases social media platforms’ rhetorical work to position AI as a technological fix solution to their own problems.

Platform companies such as Facebook, Twitter and Google, but also Uber and AirBnB, had long been very successful in positioning themselves as neutral intermediaries (Gillespie, 2010). The absolute enabling of free expression and the “information-wants-to-be-free” mantra had constituted the central guiding principles of Silicon Valley companies (Vaidhyanathan, 2012). Whether it was search results, news feeds or simply providing the opportunity to express oneself publicly: Until well into the 2010s, platform companies portrayed their services as value-free offerings and positioned themselves as tech companies, not media organizations (Napoli and Caplan, 2017). The algorithms used to sort and prioritize content would deliver objective and thus neutral results (Ames, 2018). This “spiritual deferral to algorithmic neutrality” (Morozov, 2011) was expressed, for example, in Google’s systematic refusal to take responsibility for search result lists and to change them manually, even if they showed racist or discriminatory content as the top search results (Gillespie, 2014: 180-181; Noble, 2018). Until 2015, Twitter made it clear in the very first words of its “Twitter Rules”, which are both a self-portrayal and a set of rules, that users are responsible for their own content and that Twitter remains neutral as a

provider: “We respect the ownership of the content that users share and each user is responsible for the content he or she provides. Because of these principles, we do not actively monitor and will not censor user content”.<sup>2</sup> This positioning as neutral intermediaries was highly attractive for platform providers, as they could thus establish themselves as a self-evident and soon seemingly indispensable element of everyday communication, but at the same time could neither be held socially responsible nor legally liable for the content circulating across their services. “They do so strategically, to position themselves both to pursue current and future profits, to strike a regulatory sweet spot between legislative protections that benefit them and obligations that do not, and to lay out a cultural imaginary within which their service makes sense” (Gillespie, 2010: 348). This positioning was a major factor in the rapid growth of platforms into central institutions of social communication.

However, at the latest since the 2016 US elections, and the increasing political and social conflicts on migration issues since 2015, a turn to responsibility in the debate on platforms can be observed. In these years, controversies about the role and adequate regulation of platforms have increased significantly (Katzenbach, 2021). In particular, the intense debates around misinformation or “fake news” (Righetti, 2021) and hate speech (Tworek, 2021) have placed questions of the power and responsibility of platforms high on public and political agendas. Providers are now perceived in the discourse first and foremost as active actors who organize their services in specific ways, pursuing their own interests (quantitative optimization on interactions, monetization through advertising) as well as generating external effects (reinforcing dynamics of misinformation, hate comments). In the meantime, platform providers have adapted noticeably to this “responsibility turn” by admitting mistakes, accepting responsibility and – for example Facebook – interpreting their mission of connecting all people more qualitatively than quantitatively in external communication (Lischka, 2019; Haupt, 2021). Twitter also changed their opening statement significantly. Since 2019 it reads: “Twitter’s purpose is to serve the public conversation. Violence, harassment and other similar types of behaviour discourage people from expressing themselves, and ultimately diminish the value of global public conversation. Our rules are to ensure all people can participate in the public conversation freely and safely.”<sup>3</sup> With its stark contrast to the original wording, these words clearly illustrate the responsibility turn in the public understanding of platforms. Twitter acknowledges the responsibility for the content on its service, and justifies restrictions on freedom of expression.

So while there is now a broad consensus that platforms have responsibility for the content and communication dynamics on their services, the shape of this responsibility is by no means clear. How and according to which criteria should platforms judge and, if necessary, block

controversial and problematic content? These complex questions are exacerbated by the massive size and content volume of these services, demanding to carrying out these often-difficult balances between freedom of expression on the one hand and, for example, personal rights on the other hand at scale, i.e. million or even billion times day by day. In this situation, with masses of potentially problematic content on the one hand and growing attribution of responsibility to platforms on the other, the technological fix now appeals to many as the only way out. Providers and regulatory actors have in recent years mutually reinforced the belief that technology, especially AI, can solve these problems of social media platforms. Not only has Facebook's CEO Zuckerberg regularly promised AI as a technological fix in hearings in North America and Europe (Katzenbach, 2019; Lischka, 2019; Russell, 2019). Yann LeCun, the world's leading AI researcher in Facebook's service, considers "fake news" or misinformation as technically solvable (Seetharaman, 2016); journalists regularly follow him in this view (cf. e.g. "Why fake news is a tech problem", Elgan 2017); and political actors, high-level courts and regulatory initiatives increasingly assume the efficient and balanced functioning of automatic filtering systems when drafting decisions and regulations, often without substantially taking into account their limitations and their side effects on freedom of expression (cf. the following section). In consequence, this twofold discursive development of a general discourse that positions AI as a solution to social problems and an increasing attribution of responsibility to social media platforms constitutes a powerful discursive turn to AI in governing communication.

## The political turn to AI – from liability privilege to the search for a new regime

At the political-regulatory level, political actors and regulatory bodies have in recent years initiated a search movement, particularly in the European context but increasingly also in the USA, that parallels the responsibility turn on the discursive layer. In their quest for translating the growing demand for responsibility into law and regulation, political actors, courts and regulatory agencies are turning to governance mechanism that hold social media platforms more and more accountable and, in some cases, liable for the content that they host – which in turn further pushes platforms to automatically screen their content for possibly infringing or violating content.

This development constitutes a remarkable move away to the paradigm that has dominated Internet regulation in the past twenty-five years. Since the late 1990s intermediaries and platforms have operated under a liability privilege and the notice-and-takedown procedure: Only when providers have knowledge of illegitimate content or unlawful

conduct on their services, they do have to take action by filtering content or blocking users' access. This paradigm has been established both in Europe with the EU E-Commerce Directive 2001 (Kuczerawy and Ausloos, 2015) and in the US in the US Digital Millennium Copyright Act (DMCA) and with even more extensive freedoms in Section 230 of the US Communication Decency Act (CDA) (Citron and Wittes, 2017; Gasser and Schulz, 2015; Holland et al., 2015) since around the turn of the millennium. In the European Union (EU), this supranational directive has been installed in national regulations. In Germany, for example, this downstream responsibility of social media providers has been expressed, for example, in the Telemedia Act (TMG).<sup>4</sup> Telemedia providers in Germany, including platforms, are legally responsible only for their "own information" they provide (Section 3, §7). Since the function of social media is usually understood in case law to mean that they – in the words of the TMG – only "transmit or [...] provide access to use" "third-party" information, the content in dispute does not fall under Section 3, §7. However, according to Section 3, §8 TMG, they are not responsible for the content they mediate and cannot be held liable for it. Only when they become aware of illegal content do they have to intervene (Section 3, §10).

For almost ten years, however, there has been a development in Europe towards a much narrower interpretation of this liability privilege and even a turning away from this paradigm. This "road to responsibilities" (Sithigh, 2020; Kuczerawy, 2019) is expressed both in court rulings and in regulatory initiatives, first in national rulings and law, but increasingly also on the European level. In Germany, for example, the Federal Supreme Court (BGH) reformulated the liability privilege in a ruling on the file hoster Rapidshare in a demanding way by imposing a "market monitoring obligation" on the provider, which obliges it to "determine, using suitably formulated search queries and [...] using web crawlers, whether there are indications of further infringing links on its service with regard to the specific works to be checked". In German legislation the development of increasingly strong joint liability of providers culminated in the Network Enforcement Act (NetzDG), which, primarily as a reaction to increased hate speech and misinformation in social media, obliges the major platforms to respond to complaints from netizens in short time windows (24 h for "obviously illegal" content, seven days for all others) and to delete content if necessary (Schulz, 2018). The new State Treaty on the Modernization of the Media Order in Germany ("Media State Treaty") concluded by the federal states at the end of 2019 also follows this route: social media providers will be integrated into the federal German system of broadcast media regulation as "media intermediaries" (in the new §2 para. 2 no. 16) once this State Treaty has been passed by the state parliaments. As a result, they will have to disclose the "central criteria of aggregation, selection and presentation of content

and their weighting" (§ 93 Transparency). In addition, they must ensure that individual content providers are not "discriminated against", i.e. the declared criteria must be applied indiscriminately to all content (§ 94 Freedom from discrimination). However, the impact of these new forms of regulation has yet to be seen. While the freedom from discrimination is a regulatory novelty, the transparency requirements and the increased joint liability are not a German peculiarity. In other countries such as France and England, legislators and regulation are also moving in this direction (Bunting, 2018, Sitingh 2020).

Institutions at the European level have also gradually moved away from the paradigm of liability privilege in court rulings of the highest instance and legislative initiatives, and have introduced stronger requirements and liability rules. The European Court of Justice (ECJ), in rulings such as that on the "right to be forgotten", has significantly increased the responsibility of platforms for the content they provide (Kuczerawy and Ausloos, 2015), and increasingly obliges providers to use automated procedures to prevent the re-publication of statements and content once notified as unlawful with the same content (Glawischnig-Piesczek vs. Facebook, ECJ, C-18/8, cf. Kuczerawy 2019). The EU Commission and the European Parliament are significantly increasing the legal co-responsibility of social media providers through the introduction of self-obligations of providers to immediately delete and block content glorifying terrorism and violence, through the adoption of the new copyright directive with significantly more far-reaching liability provisions (Directive EU 2019/790) and in the planned comprehensive EU Digital Services Act. In the US, regulatory initiatives to revise both the DMCA and the CDA also point towards this road to responsibility and more procedural accountability (Keller, 2020), but have not yet gained enough political momentum and constitutional consistence (Keller, 2021).

These various measures at the political-regulatory level constitute a search movement aimed at translating the consensual demand for more responsibility of platforms into adequate regulatory measures. We can clearly observe a gradual departure from the pure paradigm of liability privilege. The destination of this new route is not yet foreseeable. A switch to the antithesis, the liability of platforms for all content they convey, appears neither desirable nor realistic. The encroachment on freedom of opinion and information and the loss of diverse public spheres would be too serious (Schulz, 2019). What is already clearly foreseeable, however, is the noticeably narrower and much more demanding conditions of the liability privilege, which at least require proactive measures from providers to block illegal or infringing content that is already known. Added to this are the increasing demands to effectively detect and combat even difficult-to-classify issues such as copyright infringement, hate speech or misinformation increase (Bloch-Wehba, 2020).

These growing political-regulatory demands on platforms massively favour the algorithmic turn presented here. For how can the categorization and, if necessary, filtering of content be achieved at scale? In view of the scale of content and the size of the problems, automated procedures alone seem to be able to help. In this way, the current political-regulatory development is strongly pushing forcing the algorithmic turn in the governance of platforms.

## Contributions to this special section

Thus, there is more to the turn to AI in platform governance and the regulation of digital communication than pure technological progress. It is an entanglement of discursive and political developments, the existing technological advancements and strong economic interests. Against this backdrop, this special theme features a diverse set of articles that addresses this multi-layered turn to AI from various perspectives.

The first two articles approach the theme by investigating public discourses and expectations in the context of AI. In their piece, *Jonathan Roberge, Marius Senneville and Kevin Morin* mobilize Callon's concept of "translation" (Callon, 1986) to surface the rhetorical and discursive work that tech entrepreneurs and policy makers employ in order to position AI as an essential and indispensable element of contemporary societies. Despite substantial criticism, these actors successfully translate very different, often long-existing, sometimes buggy technologies into AI as a coherent rhetorical device that promises a better tomorrow, as the authors can show in their case studies on the Montreal Declaration for a Responsible Development of AI, the Zuckerberg Hearing at the US Senate and the integration of armed drones and robots into the military. *Aphra Kerr, Marguerite Barry and John D Kelleher* approach the theme of public discourses and expectation by focusing on the emerging debate on AI and ethics. In their article, the authors ask how societal expectations on an emerging technology such as AI are structured and how this in turn informs the further development of the phenomenon. Their study combines an analysis of documents published by key actors in AI research and policy and a public survey on awareness of AI, expectations and ethical considerations. The findings show that a range of actors construct the expectation that AI brings about massive economic opportunities but also ethical concerns. In many ways, expectations take a performative function here with actors "talking AI into being" (Bareis and Katzenbach, 2021) but saying little about the challenges of applying these technologies in particular social contexts. In consequence, the authors call for a less technology-oriented and much more situated and practice-based approach to the integration of AI into society, including more domain appropriate AI tools, updated professional practices, dignified places of work and robust regulatory and accountability frameworks.

The second focus area of this special theme focuses on such a specific domain: the automated moderation and regulation content on social media platforms. *Robert Gorwa, Reuben Binns and Christian Katzenbach* give an overview on the technical foundations and the practical deployment of automated tools in the moderation of content on social media platforms. The authors offer a typology that differentiates matching and classification as key technological logics and different types of regulations such as blocking, (de-)monetization, downranking and flagging. Social media platforms have long used such systems for identifying possible copyright infringement, they now increasingly screen their content automatically as well for hate speech, misinformation, terrorist propaganda and other harmful or controversial content. While many take issue with the often erroneous outcome of these systems, the authors make a much more fundamental argument: Even if or when such systems operate perfectly from a technical perspective, three highly political issues will always exacerbate rather than relieve: algorithmic content moderation threatens to further increase opacity, to further complicate outstanding issues of fairness and justice and to re-obscure the fundamentally political nature of speech decisions being executed at scale. With these high issues at stake, *Joanne E Gray and Nicolas P Suzor* turn themselves to AI technologies to understand the social media platforms' automated content moderation practices. With a methodological experimentation built on machine learning technologies, they investigate a dataset of almost 80 million YouTube for patterns in removal of content. In the substantial dimension, the article shows that content is mostly blocked because of user account termination (4%). 0.77% of content is being blocked because YouTube's Content ID automatically flagged the content as copyright infringement; only 0.57% is taken due to infringements of YouTube's Terms of Service and 0.11% because of copyright owners' DMCA requests. On the methodological dimension, the article offers a fruitful approach to not only investigate content moderation at scale but to proper understand at scale content moderation.

The third and last part of the special theme offers analyses of the political ramifications and challenges of the turn to AI in communication governance. In her commentary, *Emma J Llansó* argues that despite the current automated measures no amount of AI in content moderation will solve filtering's prior-restraint problem. Llansó reminds us that the legal concept of prior restraint – that a speaker must seek approval from some empowered third party – is typically considered to be incompatible with international human rights law. The current policy and technological turn to AI and proactive measures in content moderation is exposing more and more speech online to evaluation and approval, and thus constitutes not only a road to responsibility but also a return trip to prior restraint. In consequence, Llansó calls on governments and regulators to strictly restrain from filtering mandates, and

on companies to recognize the human rights risks inherent in their content moderation systems and to really thrive to mitigate them. *Niva Elkin-Koren* approaches these challenges from a different angle in her research article. The author agrees with Llansó's analysis that filters carry censorial power, potentially bypassing traditional checks and balances secured by public law and easily escaping oversight due to their opaque and dynamic nature. Her suggestion, though, is to counter these challenges with rather more than less AI: with public interest driven AI technologies that operate with an adversarial logic. Such an initiative could counterbalance the single optimization standard of current content removal systems by introducing other trained AI systems that follow different logics such as pluralism or existing case law. The concluding commentary by *Tarleton Gillespie* finally articulates the elephant in the room: the issue of scale. The turn to AI is often justified as a response to scale: social media platforms have grown so large that only automated measures seem to be able to handle the massive amounts of content, and that AI systems thus appear desirable, even inevitable. At the same time, as the articles of this special theme have shown, a great many problems remain – even with highly optimized and efficient systems. But what is the consequence then? Maybe, if moderation is so challenging at scale, Gillespie argues, we should conceive of this as a limiting factor on the “growth at all costs” mentality.

With these different perspectives on the multi-layered turn to AI, the articles of this special theme may not always offer clear answers but they articulate the stakes of the current development and raise the key questions. Given the already deep integration of these technologies into our daily lives (Hepp, 2020) and the massive economic power of the major companies, the medium- and long-term challenge will be to continually establish and address these fundamental political questions on the public and political agenda. The history of science and technology teaches us that once a technological fix is successfully installed – i.e., established as successful, regardless of its actual effects – such questions tend to disappear from the space of debate and negotiation, and become infrastructures that are taken for granted. We can already anticipate that algorithmic moderation and regulation will become more and more seamlessly integrated into our social lives. We hope that with this special theme, we contribute to counteract such normalization of automated decisions and to keep questions of content moderation and the automation of the social on the public and political agenda – because we know at least one thing: AI will never fix this, whatever this is.


### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

Research underlying this article has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement number 870626 ("ReCreating Europe") and from the German Research Foundation (DFG) as part of the multinational scheme "Open Research Area" (ORA) under grant number 440899634.

## ORCID iD

Christian Katzenbach  <https://orcid.org/0000-0003-1897-2783>

## Notes

1. Cf. the video and official documentation of the hearing on the US Senate website (<https://www.judiciary.senate.gov/meetings/facebook-social-media-privacy-and-the-use-and-abuse-of-data>), as well as the verbatim transcript at: [https://en.wikisource.org/wiki/Zuckerberg\\_Senate\\_Transcript\\_2018](https://en.wikisource.org/wiki/Zuckerberg_Senate_Transcript_2018).
2. Cf. Twitter Rules in the version of 2009 as archived by the *Platform Governance Archive* (Katzenbach et al. 2021), available at: <https://pga.hiig.de/view/twitter/cg/2009-01-18>, ll. 4–8.
3. Cf. Twitter Rules in the version of 2019 as archived by the *Platform Governance Archive* (Katzenbach et al. 2021), available at <https://pga.hiig.de/view/twitter/cg/2019-06-07>, ll. 2–5.
4. Available at: <https://dejure.org/gesetze/TMG>

## References

- Ames MG (2018) Deconstructing the algorithmic sublime. *Big Data & Society* 5(1): 1–4.
- Bareis J and Katzenbach C (2021) Talking AI into being: The narratives and imaginaries of national AI strategies and their performative politics. *Science, Technology, & Human Values*: 1–27. <https://doi.org/10.1177/01622439211030007>.
- Bloch-Wehba H (2020) Automation in moderation. *Cornell International Law Journal* 53: 41–96.
- Brennen JS, Howard PN and Nielsen RK (2018) An industry-Led debate: How UK Media cover artificial intelligence. *Reuters Institute for the Study of Journalism*: 10.
- Bory P (2019) Deep new: The shifting narratives of artificial intelligence from deep blue to AlphaGo. *Convergence*: 1–16. <https://doi.org/10.1177/1354856519829679>.
- Bunting M (2018) From editorial obligation to procedural accountability: Policy approaches to online content in the era of information intermediaries. *Journal of Cyber Policy* 3(2): 165–186.
- Callon M (1986) Some elements of a sociology of translation: Domestication of the scallops and the fishermen of St brieuc Bay. In: Law J (eds) *Power, Action, and Belief: A New Sociology of Knowledge?* London: Routledge and Kegan, 196–233.
- Campolo A and Crawford K (2020) Enchanted determinism: Power without responsibility in artificial intelligence. *Engaging Science, Technology, and Society* 6: 1.
- Cardon D, Cointet J-P and Mazières A (2018) Neurons spike back: The invention of inductive machines and the artificial intelligence controversy. *Réseaux* 211(5): 173.
- Cave S and Dihal K (2019) Hopes and fears for intelligent machines in fiction and reality. *Nature Machine Intelligence* 1(2): 74.
- Chuan C-H, Tsai W-HS and Cho SY (2019) Framing Artificial Intelligence in American Newspapers. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society - AIES '19, pp.339–344. <https://doi.org/10.1145/3306618.3314285>.
- Citron DK and Wittes B (2017) The internet will Not break: Denying Bad samaritans section 230 immunity. *Fordham Law Review* 86(2): 401–423.
- Daub A (2020) *What Tech Calls Thinking: An Inquiry Into the Intellectual Bedrock of Silicon Valley*. New York: Farrar Strauss & Giroux.
- Degele N (2002) *Einführung in die Techniksoziologie*. München: Fink.
- Elgan M (2017) why fake news is a tech problem. With fake news wrecking everything, silicon valley is our last hope. *Computer World*. <https://www.computerworld.com/article/3162020/why-fake-news-is-a-tech-problem.html>.
- Fast E and Horvitz E (2017) Long-Term Trends in the Public Perception of Artificial Intelligence. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17).
- Fischer S and Puschmann C (2021) *Wie Deutschland über Algorithmen Schreibt. Eine Analyse des Mediendiskurses über Algorithmen und Künstliche Intelligenz (2005–2020)*. Gütersloh: Bertelsmann Stiftung. <https://doi.org/10.11586/2021003>.
- Gasser U and Schulz W (2015) Governance of online intermediaries: Observations from a series of national case studies. *Berkman Center Research Publication* 2015(5): 1–27. <https://doi.org/10.2139/ssrn.2566364>.
- Gillespie T (2010) The politics of 'platforms'. *New Media & Society* 12(3): 347–364.
- Gillespie T (2014) The relevance of algorithms. In: Gillespie T, Boczkowski P and Foot K (eds) *Media Technologies*. Cambridge, MA: MIT Press, 167–193.
- Gorwa R, Binns R and Katzenbach C (2020) Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society* 7(1): 1–15.
- Haupt J (2021) Facebook futures: Mark Zuckerberg's Discursive construction of a better world. *New Media & Society* 23(2): 237–257.
- Hepp A (2020) *Deep Mediatization*. London: Routledge.
- Holland A, Bavitz C, Hermes J, et al. (2015) *NoC Online Intermediaries Case Studies Series: Intermediary Liability in the United States*. 70. Berkman Centre for Internet & Society. Cambridge, MA. <https://perma.cc/2QAY-UTDY>.
- Katzenbach C (2019) AI will fix this. In: Kettemann M and Dreyer S (eds) *Busted! The Truth About the 50 Most Common Internet Myths*. Internet Governance Forum, Berlin: Leibniz Institute for Media Research | Hans-Bredow-Institute, 194–195. [https://www.hiig.de/wp-content/uploads/2019/11/dgzmogf\\_KettemannDreyerInternetMyths2019-1.pdf](https://www.hiig.de/wp-content/uploads/2019/11/dgzmogf_KettemannDreyerInternetMyths2019-1.pdf).
- Katzenbach C (2021) Die Öffentlichkeit der plattformen: Wechselseitige (Re-)institutionalisierung von Öffentlichkeiten und plattformen. In: Eisenegger M, Prinzing M, Ettinger P and Blum R (eds) *Digitaler Strukturwandel der Öffentlichkeit*. Springer Fachmedien Wiesbaden, 65–80.
- Katzenbach C, Magalhães JC, Kopps A, et al. (2021) The Platform Governance Archive (PGA). Berlin. <https://doi.org/10.17605/OSF.IO/XSBPT>.

- Keller D (2020) *CDA 230 Reform Grows Up: The Pact Act Has Problems, But It's Talking About The Right Things*. Stanford: Center for Internet and Society. <http://cyberlaw.stanford.edu/blog/2020/07/cda-230-reform-grows-pact-act-has-problems-it-s-talking-about-right-things>.
- Keller D (2021) *Six Constitutional Hurdles For Platform Speech Regulation*. Stanford: Center for Internet and Society. <http://cyberlaw.stanford.edu/blog/2021/01/six-constitutional-hurdles-platform-speech-regulation-0>.
- Kuczerawy A (2019) General monitoring obligations: A New cornerstone of internet regulation in the EU? (SSRN Scholarly Paper ID 3449170). *Social Science Research Network*. <https://papers.ssrn.com/abstract=3449170>.
- Kuczerawy A and Ausloos J (2015) From notice-and-takedown to notice-and-delist: Implementing google Spain. *Colorado Technology Law Journal* 14(2): 219–258.
- Lischka JA (2019) Strategic communication as discursive institutional work: A critical discourse analysis of mark Zuckerberg's Legitimacy talk at the european parliament. *International Journal of Strategic Communication* 13(3): 197–213.
- Morozov E (2011) Don't be evil. *The New Republic*. <http://www.newrepublic.com/article/books/magazine/91916/google-schmidt-obama-gates-technocrats>.
- Morozov E (2013) *To Save Everything, Click Here: The Folly of Technological Solutionism*. New York: PublicAffairs.
- Napoli P and Caplan R (2017) Why media companies insist they're not media companies, why they're wrong, and why it matters. *First Monday* 22(5): 1–16. <https://doi.org/10.5210/fm.v22i5.7051>.
- Noble SU (2018) *Algorithms of Oppression: How Search Engines Reinforce Racism*, 1 edition New York: NYU Press.
- Righetti N (2021) Four years of fake news. *A Quantitative Analysis of the Scientific Literature [Preprint]*: 1–24. SocArXiv. <https://doi.org/10.31235/osf.io/buemr>.
- Russell FM (2019) The New gatekeepers. *Journalism Studies* 20(5): 631–648.
- Seetharaman D (2016) Facebook looks to harness artificial intelligence to weed Out fake news. Company executives say the social network first needs a policy on how to responsibly apply such capabilities. *Wall Street Journal*. [https://www.wsj.com/articles/facebook-could-develop-artificial-intelligence-to-weed-out-fake-news-1480608004?mod=pls\\_whats\\_news\\_us\\_business\\_f](https://www.wsj.com/articles/facebook-could-develop-artificial-intelligence-to-weed-out-fake-news-1480608004?mod=pls_whats_news_us_business_f)
- Schulz W (2018) Regulating Intermediaries to Protect Privacy Online – the Case of the German NetzDG. HIIG Discussion Paper Series, 2018–01. <https://www.hiig.de/wp-content/uploads/2018/07/SSRN-id3216572.pdf>.
- Schulz W (2019) Roles and Responsibilities of Information Intermediaries. Fighting Misinformation as a Test Case for a Human Rights-Respecting Governance of Social Media Platforms. Hoover Institution, Stanford University, USA: Aegis series Paper 1904.
- Sithigh DM (2020) The road to responsibilities: New attitudes towards internet intermediaries. *Information & Communications Technology Law* 29(1): 1–21.
- Tworek HJS (2021) Fighting hate with speech Law: Media and German visions of democracy. *The Journal of Holocaust Research* 35(2): 106–122.
- Vaidhyanathan S (2012) *The Googlization of Everything (And Why We Should Worry)*. Berkeley: University of California Press.
- Volti R (2014) *Society and Technological Change*. 1995. Duffield: Worth Publishers.
- Zeng J, Chan C and Schäfer MS (2020) Contested Chinese dreams of AI? Public discourse about artificial intelligence on WeChat and People's Daily online. *Information, Communication & Society*: 1–22. <https://doi.org/10.1080/1369118X.2020.1776372>.