Global Media Journal
Journal for International and Transcultural Communication
German Edition

BerlinUP Journals

Special Section: Norms, Power Relations and Injustices in Digitality
Editorial Essay

Norms, Power Relations and Injustices in Digitality: Global Perspectives. An Introduction to the Special Section on Content Moderation

Christoph Böhm & Oliver Zöllner

Abstract: This introduction to the Global Media Journal – German Edition's Special Section on content moderation critically examines the governance challenges of regulating online content in the digital age. It argues that the removal of content deemed illicit, detrimental or otherwise unacceptable by established standards is often framed as neutral gatekeeping even though it operates within contested power dynamics that lack rigorous and sound frameworks to assess its real-world impacts on freedom of expression, safety, and social justice. The authors highlight historical parallels to premodern censorship struggles following the invention of the printing press, emphasising that digital public spheres require forms of robust civility, not mere technical fixes. Current practices of content moderation reveal deep tensions: platforms increasingly employ automated systems for this endeavour, yet this risks amplifying harmful content and creating ethical dilemmas, while low-wage, high-stress labour conditions for human moderators ("clickworkers") expose systemic exploitation. The Special Section addresses these gaps through four case studies.

Keywords: content moderation, gatekeeping, social media, digital public sphere, freedom of expression, censorship

Author information:
Christoph Böhm (Dr. phil.) is a computer scientist and worked with the SAP company until 2024, with key responsibilities including corporate strategy and business ethics. After gaining a doctoral degree in philosophy from the University of Freiburg in 2023, he has pursued independent research in the field of responsible digitalisation. In his most recent professional activities, Böhm focused on the sustainability of information technology. The central themes of his current scholarly work involve the practical application and scholarly communication of philosophical insights into digitality. One of his primary objectives is to promote responsible engagement with digital technologies across business, politics, and society.
For more information: https://infosophia.de
ORCID: 0009-0000-6294-1699
E-Mail: christoph.boehm@infosophia.de

Global Media Journal – German Edition
Vol. 15 No. 2 2025
ISSN 2196-4807
DOI: 10.60678/gmj-de.v15i2.360

Oliver Zöllner (Prof. Dr. phil.) is professor of media research, media sociology and digital ethics in the Digital and Media Economics programme at Stuttgart Media University and also teaches as an honorary professor at Heinrich Heine University Düsseldorf, Germany. He is co-founder and joint director of the Institute for Digital Ethics at Stuttgart Media University. Zöllner's current research interests include questions of digital transformation and related aspects of reflexive media literacy and digital ethics. From 1997 to 2004 he was the head of the market and media research department of Deutsche Welle, Germany's international broadcaster.
ORCID: 0000-0001-9125-1048
E-Mail: zoellner@hdm-stuttgart.de

To cite this article: Böhm, Christoph & Zöllner, Oliver (2025). Norms, Power Relations and Injustices in Digitality: Global Perspectives. An Introduction to the Special Section on Content Moderation. *Global Media Journal – German Edition*, *15*(2), DOI: 10.60678/gmj-de.v15i2.360

With the arrival of social networking sites and other platforms of user-generated content in the 2000s, online content moderation has emerged as a critical governance challenge for the entire web infrastructure. It is shaping the boundaries of the freedom of expression, safety, and community standards across global platforms, and mediating public discourse, social interaction, and information dissemination (Badouard & Bellon, 2025; Oliva, 2020). Despite its growing importance, content moderation is based on largely contested rules for the new kind of gatekeeping it seeks to do and lacks robust frameworks for analysing how different governance models perform when put to everyday practice (see Schroeder, 2025).

The special section "Norms, Power Relations and Injustices in Digitality: Global Perspectives" of this issue of the Global Media Journal – German Edition seeks to provide insights into relevant concepts, actors, and challenges in online content moderation. The case studies presented here examine the technological infrastructure underpinning moderation systems, the organisational and ethical dimensions of decision-making, and the evolving legal and policy landscapes. By synthesising current research and identifying critical research gaps, this special section aims to equip readers with the basic conceptual tools for critically engaging with this rapidly evolving domain.

The Problem

It is not only hate speech and 'shitstorms' on the internet that are becoming a problem. Live-streamed suicides and depictions of child rape in digital media are becoming increasingly widespread. Growing problematic online content is therefore prompting society and political institutions to seek legal solutions to ban toxic content from digital networks. On a metaphorical level, this endeavour could be compared to efforts to remove the ever-increasing amount of plastic waste from the world's oceans, to cite a well-worn, but fitting image: it is obviously a task of Sisyphean scope. First official reports that addressed the idea of sanctions for criminal online content were published in the early 2010s (see European Union Agency for Fundamental Rights, 2012), later to resurface most prominently in the EU's Digital Services Act (DSA) with its out-of-court dispute settlement bodies, and other detailed suggestions for the removal of undesirable content (European Union Agency for Fundamental Rights, 2023; see Kira, 2025). Since then, sophisticated regulations around the world have required social media platform operators to remove undesirable content from the internet as quickly as possible after its publication. In response to external pressure, tech companies and their outlets began to set up their own structures for identifying and deleting offensive posts. In doing so, they laid the foundation for the widespread practice of *content moderation*.

However, this term may be deemed a euphemism. Behind the terminology lies a practice of deletion (amounting to, in the eyes of some, de-facto censorship), and not, as one might assume, a well-balanced negotiation process in which different interests are reconciled, e.g., through the mediation of a neutral authority. But how exactly, by whom, and under whose legislation, jurisdiction and legally binding oversight? Regulatory answers to these urgent questions remain all too often vague. Due to the growing number and ever-proliferating intensity of disturbing content on social media, e-commerce platforms, video-sharing, and messaging services alike, and despite patchy legal frameworks, content moderation has nonetheless developed into an industry in its own right, even though the data centres where this work is carried out are largely invisible to the general public and its awareness.

Is this phenomenon of feuds over what can be said really new? Jürgen Habermas (2022) observes a fundamental change in digital media in that free platform access means that everyone can become a potential author or publisher – in fact, this was once one of the key utopian visions for establishing what eventually became the internet. Indeed, the former role of the media as gatekeepers has disappeared with the advent of unedited distribution channels (Habermas, 2022, p. 46). Traditional publishers and media institutions implicitly assume this controlling role by being bound to the morality of their respective society or target audience through their editorial policies, or programmatic orientation, on the one hand and generally accepted media standards on the other.

However, the internet and, above all, social media platforms are free from content-related programmatic considerations and function purely as distribution platforms. Nevertheless, publications are subject to intervention, and most often so for economic reasons. Content that has the potential to trigger high levels of follow-up activity in digital media through forwarding or other practices of user engagement is algorithmically amplified and rewarded with higher visibility (see Aral, 2020, pp. 56–92; Zuboff, 2019, pp. 199–232). In the context of this phenomenon, Habermas asks the pointed question of how many centuries it took for people to *learn to read* and how long it might now take for them to *learn to write* in the sense of being free authors in the, historically speaking, still relatively new kinds of digital environments.

Historic Antecedents

In a certain way this analogy is correct, yet it does not address the fundamental problem of digital attention production. It is true that the invention of the printing press in Europe (in ca. 1440) has had a decisive influence on the course of history. Among other things, the mechanised reproduction and dissemination of writings played a prominent role in the interpretation and spreading of the Christian faith. In the growing controversy over indulgences of the clergy, the theologian Martin Luther was able to initiate the Protestant movement against a Roman Catholic church riddled with kleptocracy. Using the technology of printing, in 1517 Luther published a protest statement comprising of 95 theses which revolutionised the religious beliefs of the time. Numerous counter-publications and, eventually, even religiously motivated warfare resulted (see Puchner, 2017, pp. 145–169).

For our discussion it is important to remind ourselves of the fact that the widespread unrest sparked by the free availability of printed texts and the disruption it caused – the "written world" that came into being as a result of a technological revolution (Puchner, 2017; see also McLuhan, 1962) – motivated ecclesiastical and monarchical authorities to resort to censorship, not least in order to maintain the order of state and church – an effort that initially, however, proved ineffective. Over time, local censorship rules emerged in the fragmented principalities and kingdoms of Europe. It was not until the beginning of Enlightenment in the 18th century that new understandings of freedom linked to the concept of *Bildung* (education imbued with individual development) emerged, which ultimately laid the foundation for the fundamental right to freedom of expression after the Second World War (see Andersen & Björkman, 2024).

Thus, several hundred years after the invention of the printing press, this part of *learning to read* was completed as a social process. In this sense Habermas' analogy may be considered accurate. However, against the background of the historical antecedents described above, the idea of entering a new age of social conflict over what other people have communicated online just for society to learn free authorship would seem inappropriate and should, from a perspective of ethics, be averted at all costs. For free citizens to communicate freely in digital environments entails the need for society to provide for an infrastructure to eliminate or at least moderate what is deemed undesirable for the sake of sustainable free speech and free societies oriented towards a flourishing life for their citizens.

For this end, "not just civility but robust civility" is needed (see Garton Ash, 2016, p. 212). Content moderation means to indeed delete certain content agreed to be undesirable or harmful, but in a Western liberal democracy this form of moderation should not be regarded as censorship. However, recent tendencies in some liberal democracies and their technology sectors indicate that such a form of regulation and oversight is anathema to some political and entrepreneurial protagonists and their libertarian – or "cyberlibertarian" – worldviews (see Golumbia, 2024, pp. 302–325) that seem to focus on data extraction, corporate power, and behavioural manipulation via online platforms. It is obvious that civil society should voice its concerns and steer debates and policies in directions that are in line with the values of liberal democracy (see Michalon, 2025; Palladino et al., 2025).

Current Challenges of Content Moderation

In digital media, social mobilisation by way of algorithmically enhanced user engagement ("infinite scrolling"; Tortorici, 2020) seems to be a major goal of blatant deviations from traditional journalistic and editorial policies as practiced by 'legacy' media in the context of liberal media systems. Such deviations from well-established norms of both professional and social conduct increasingly seem to make themselves heard ever more loudly against the background noise of digital media streams. In this respect, it seems advisable at this point to take a close look at the media content that can trigger social disruption and shock. Problematic, disturbing and sometimes unbearable content reveals hidden social conflicts that actually require a multi-layered governance approach based on accountability in the digital realm (see Clune & McDaid, 2024). Whether content moderation may, under certain circumstances and for certain target groups, "do more harm than good", e.g., if exaggerated or overreaching, is a debatable point (see Zhang et al., 2024, in the context of juvenile mental health). What emerges is a need for forms of content moderation that take on functions of organising and mediating processes for systematically and holistically tackling issues of violations of dignity, rights (see Oehmer-Pedrazzi & Pedrazzi, 2025) and, more broadly, of an orientation towards the truth in digital environments.

Instead, the companies owning social networking platforms are increasingly using applications of artificial intelligence (AI) to automatically manage the flood of harmful content on their platforms (see Karabulut et al., 2023, for an experimental algorithmic case study). This technology-based form of content moderation may lead to a situation in which not less, but much more problematic online content is created and disseminated, namely by the artificial systems themselves. Digital corporations are well aware of this. As Jörg (2024) points out, "incorrectly coded moderation algorithms might undermine the 'epistemic potential' of political discourse" (p. 247). Therefore, additional levels of nuanced monitoring are required for digital content that cannot be clearly identified by AI and is in need of human decision-making (see Gongane et al., 2022; Wiesner et al., 2025).

To this end, people are employed in low-wage locations, often in precarious working conditions amounting to exploitation. Under conditions of severe psychological stress, employees are instructed to view web content that has been marked as possibly problematic, and decide in a matter of seconds whether it should be deleted or not (see Grassegger & Krause, 2016). This practice of *cleaning up the internet*, shared between humans and machines, creates an ethical dilemma in which humans become the guardians of problematic content that is swelling due to technological amplification. However, this idea or figure of thought is not entirely new either. In the wake of increasing technological advancement and the arrival of new technology seemingly beyond the immediate control of individuals, the philosopher Günther Anders (1980) diagnosed that humans are increasingly taking on the role of "shepherds" of the technology that they themselves have created. This 'shepherding' is linked to a loss of humans' mastery of technology, i.e., their superior position when it comes to using and handling technological systems. Humans are thus reduced to mere administrators or "guardians" that "maintain" or "wait on" these systems of their "product-and gadget-world" (Anders, 1980, p. 281; trans.).

Looking at today's guardians of digital networks and environments, all too often ill-paid clickworkers act as content moderators operating under frequently precarious and exploitative labour conditions. These clickworkers, "guardians" of concurrent digital technologies, perform their tasks in corporate or outsourced data centres under enormous pressure and strain (see Grassegger & Krause, 2016). Under these conditions, this task – present-day content moderation in the real world – is manageable neither psychologically nor materially, and produces harm (Spence et al., 2023). Therefore, the debate about how to deal with content moderation appropriately must not be left to the platforms themselves. The enlightened self that has emerged from crises, conflicts and wars has painfully acquired the tools to overcome problematic developments in society by naming abuses and engaging in discourse aimed at overcoming them. Debates must be held. This special section contributes to this with the following case studies.

Articles featured in this special sections

Lukas Beckmann, Sebastian Suttner and *Björn Wiegärtner* examine content moderation from a systems theory perspective as a form of communication control. Rather than arriving at normative or legal classifications, they understand content moderation as a practical problem within communicative systems and introduce three historical social figures of thought – "chaperone", "referee" and "censor" – that serve as heuristic points of comparison to illustrate different modes and problems of communication control: from interactive monitoring to organisational rule enforcement to social control of what can be said. Their article concludes that content moderation should be understood as a practice whose conflicting goals cannot be conclusively resolved, but whose management itself becomes a permanent social task.

Sarah Rebecca Strömel and *Lea Watzinger* look at "Transparency and Deplatforming as Strategies of Debate in Digital Public Spaces". Their contribution explores how the principle of deplatforming may enable the exclusion of radical, dehumanising actors and positions from digital spaces, thereby denying them access to platforms and public visibility. The opposing principle of transparency aims to represent diverse viewpoints in their full range, including counter-arguments and dissenting perspectives. Both approaches shape the public sphere and may be guided by differing normative ideals of publicness. The authors elaborate on these two principles and situate them within democratic theory: by embedding them in deliberative and radical democratic perspectives on the public sphere and its digital transformation, they examine the underlying democratic assumptions that are implicitly made when choosing deplatforming or transparency as regulatory strategies.

Mariana Magalhães Avelar and *Luana Mathias Souto* focus on the perspective of "Genderwashing by Digital Platforms' Self-regulations" and analyse decisions taken by the Meta Oversight Board, an independent body that evaluates Meta Inc.'s decisions on content removal or moderation across Facebook, Instagram, and Threads. The article examines two binding decisions and asks what legal values, norms, and principles were applied in these gender-sensitive rulings, and whether these decisions led to real protection, meaningful remedies, or changes in Meta's platform infrastructure. The authors argue that Meta's responses are largely superficial, using human-rights language to appear accountable, but failing to tackle the root problem: the platform's underlying design and infrastructure. As a result, many of these actions function more as symbolic gestures than as genuine, systemic reforms.

Jonathan D. Geiger's article "Structuring the Infosphere Online" examines web search engines as infrastructure of content moderation. These engines help users find relevant information by organising and structuring the web's content. The article looks at how search engine results are generated, from crawling the web to delivering results to users, and subsequently highlights where automated algorithms and human moderators are involved in shaping what appears in search results. The case study shows that these processes do not just create technical biases, but also reflect deliberate political and corporate decisions that influence what information users see. In their capacity as gatekeepers, search engines play a crucial role in shaping access to information, making them central nodes in the digital information ecosystem, Geiger concludes.

References

Anders, G. (1980). *Die Antiquiertheit des Menschen* (Bd. 2: *Über die Zerstörung des Lebens im Zeitalter der dritten industriellen Revolution*). C. H. Beck.

Aral, S. (2020). *The hype machine: How social media disrupts our elections, our economy, and our health—and how we must adapt*. Currency.

Andersen, L. R., & Björkman, T. (2024). *The Nordic secret: A European story of beauty and freedom* (2nd ed.). Nordic Bildung.

Badouard, R., & Bellon, A. (2025). Introduction to the special issue on content moderation on digital platforms. *Internet Policy Review, 14*(1). https://doi.org/10.14763/2025.1.2005

Clune, C., & McDaid, E. (2024). Content moderation on social media: Constructing accountability in the digital space. *Accounting, Auditing & Accountability Journal, 37*(1), 257–279. https://doi.org/10.1108/AAAJ-11-2022-6119

European Union Agency for Fundamental Rights. (2012). *Making hate crime visible in the European Union: Acknowledging victims' rights*. https://fra.europa.eu/sites/default/files/fra-2012_hate-crime.pdf

European Union Agency for Fundamental Rights. (2023). *Online content moderation: Current challenges in detecting hate speech*. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2023-online-content-moderation_en.pdf

Garton Ash, T. (2016). *Free speech: Ten principles for a connected world*. Atlantic Books.

Golumbia, D. (2024). *Cyberlibertarianism: The right-wing politics of digital technology*. University of Minnesota Press.

Gongane, V. U., Munot, M. V., & Anuse, A. D. (2022). Detection and moderation of detrimental content on social media platforms: Current status and future directions. *Social Network Analysis and Mining, 12*, Article 129. https://doi.org/10.1007/s13278-022-00951-3

Grassegger, H., & Krause, T. (2016). Im Netz des Bösen. *Süddeutsche Zeitung Magazin*, (50), 14–23.

Habermas, J. (2022). *Ein neuer Strukturwandel der Öffentlichkeit und die deliberative Politik*. Suhrkamp.

Jörg, S. (2024). Democratic autonomy vs. algorithms? Limits and opportunities for public reasoning. In M. Reder & C. Koska (Eds.), *Künstliche Intelligenz und ethische Verantwortung* (pp. 235–255). transcript.

Karabulut, D., Ozcinar, C., & Anbarjafari, G. (2023). Automatic content moderation on social media. *Multimedia Tools and Applications, 82*, 4439–4463. https://doi.org/10.1007/s11042-022-11968-3

Kira, B. (2025). Regulatory intermediaries in content moderation. *Internet Policy Review, 14*(1). https://doi.org/10.14763/2025.1.1824

McLuhan, M. (1962). *The Gutenberg galaxy: The making of typographic man*. University of Toronto Press.

Michalon, B. (2025). The role of civil society organisations in co-regulating online hate speech in the EU: A bounded empowerment. *Internet Policy Review, 14*(1). https://doi.org/10.14763/2025.1.1826

Oehmer-Pedrazzi, F., & Pedrazzi, S. (2025). Maßnahmen gegen Online-Hass(bilder): Zur Governance von diskriminierenden, beleidigenden oder zu Gewalt aufrufenden (visuellen) Inhalten im Netz. *Medien & Kommunikationswissenschaft, 73*(2), 235–251. https://doi.org/10.5771/1615-634X-2025-2-235

Oliva, T. D. (2020). Content moderation technologies: Applying human rights standards to protect freedom of expression. *Human Rights Law Review, 20*, 607–640. https://doi.org/10.1093/hrlr/ngaa032

Palladino, N., Redeker, D., & Celeste, E. (2025). Civil society's role in constitutionalizing global content governance. *Internet Policy Review, 14*(1). https://doi.org/10.14763/2025.1.1830

Puchner, M. (2017). *The written world: The power of stories to shape people, history, and civilization*. Random House.

Schroeder, R. (2025). Content moderation and the digital transformations of gatekeeping. *Policy & Internet, 17*, e425. https://doi.org/10.1002/poi3.425

Spence, R., Bifulco, A., Bradbury, P., Martellozzo, E., & DeMarco, J. (2023). The psychological impacts of content moderation on content moderators: A qualitative study. *Cyberpsychology, 17*(4), Article 8. https://doi.org/10.5817/CP2023-4-8

Tortorici, D. (2020, January 31). Infinite scroll: Life under Instagram. *The Guardian*. https://www.theguardian.com/technology/2020/jan/31/infinite-scroll-life-under-instagram

Wiesner, A., Schäfer, S., & Lecheler, S. (2025). Navigating the gray areas of content moderation: Professional moderators' perspectives on uncivil user comments and the role of (AI-based) technological tools. *New Media & Society, 27*(3), 1215–1234. https://doi.org/10.1177/14614448231190901

Zhang, C. C., Zaleski, G., Kailley, J. N., Teng, K. A., English, M., Riminchan, A., & Robillard, J. M. (2024). Debate: Social media content moderation may do more harm than good for youth mental health. *Child and Adolescent Mental Health, 29*(1), 104–106. https://doi.org/10.1111/camh.12689

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for the future at the new frontier of power*. Profile Books.