weizenbaum
PROCEEDINGS

Conf 2025
05

Empowering People in Online Spaces: Democracy and Well-Being in Digital Societies

Weizenbaum Conference 2025 – Book of Abstracts

Empowering People in Online Spaces: Democracy and Well-Being in Digital Societies

Weizenbaum Conference 2025 – Book of Abstracts

WEIZENBAUM PROCEEDINGS #05

BERLIN, SEPTEMBER 2025

DOI 10.34669/WI.WP/5 \ ISSN 2510-7666

Weizenbaum-Institut e.V.

Hardenbergstraße 32 \ 10623 Berlin \ Tel.: +49 30 700141-001

info@weizenbaum-institut.de \ www.weizenbaum-institut.de/en

CONFERENCE CHAIRS & VOLUME EDITORS

Clara Iglesias Keller

Jakob Ohme

EDITORIAL MANAGER

Moritz Buchner

LICENSE

WEIZENBAUM INSTITUTE

The Weizenbaum Institute is a joint project funded by the German Federal Ministry of Research, Technology and Space (BMFTR) and the State of Berlin. It conducts interdisciplinary and basic research on the digital transformation of society and provides evidence- and value-based options for action in order to shape digitalization in a sustainable, self-determined and responsible manner.

weizenbaum institut

With funding from the:

Federal Ministry of Research, Technology and Space

BERLIN

Senatsverwaltung für Wissenschaft, Gesundheit und Pflege

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Contents

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Preface

The 7th Weizenbaum Conference, entitled “Empowering People in Online Spaces: Democracy and Well-being in Digital Societies,” took place in Berlin on 5–6 June 2025. Nearly 300 participants and speakers from over 25 countries came together to explore the digital challenges facing democracies today. The conference provided a platform for interdisciplinary and international exchange, focusing on democratic participation, digital infrastructures, and well-being in the digital age. It was convened by Weizenbaum research group leaders Clara Iglesias Keller and Jakob Ohme.

The present collection of abstracts offers an overview of the talks and topics discussed at the conference. We would like to thank all participants for their valuable contributions. A total of 36 out of 58 speakers revised and submitted their long or short abstracts for inclusion.

Berlin, August 2025

Clara Iglesias Keller
Jakob Ohme
Moritz Buchner

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Panel 1

Session 1: Governance and Regulation

Session 2: Elections and Narrative Dispute

Session 3: Gen AI and Algorithmic Impact

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Comparing State-Led Content Moderation Regulations Across Nations

Augusto Santos \\ University of Erfurt \\ Erfurt \\ Germany
augusto.da_silva_santos@uni-erfurt.de

Regina Cazzamatta \\ University of Erfurt \\ Erfurt \\ Germany
regina.cazzamatta@uni-erfurt.de

Carlo José Napolitano \\ São Paulo State University \\ Bauru \\ Brazil
carlo.napolitano@unesp.br

KEYWORDS
platform regulation; misinformation; transparency; content moderation; freedom of expression

Contextual and Theoretical Background

Comparative analysis of state intervention in content moderation remains a developing field (Gorwa 2024). To contribute to the literature in platform regulation and misinformation responses, this study examines domestic regulatory approaches established in Germany (NetzDG), France (Law 2018-1202), Brazil (Resolution 23.732/2024), Singapore (Protection from Online Falsehoods and Manipulation Act, POFMA), and Turkey (Law No. 2022-7418). To varying degrees, they tackle online falsehoods and hold platforms accountable. They also illustrate either innovative domestic approaches or distinct political contexts. Our research provides a qualitative comparison of these cases from empirical (how regulations differ) and normative (how they align with supranational recommendations) perspectives. We sought to juxtapose these jurisdictions with the UN and UNESCO's guidelines on platform regulation, which are grounded in international human rights standards.

Online misinformation poses a threat, but national regulations aimed at counteracting it can also be detrimental when they violate human rights, particularly freedom of expression, speech, and access to information (Jungherr, 2024). State intervention in content moderation without alignment with international rights may lead to unwarranted censorship and disproportionately penalise platforms and users. The absence of a global regulatory framework (Gorwa, 2024)

further complicates the issue, leaving content moderation standards unclear (Gillespie, 2018). In this context, the UN (2022) and UNESCO (2023) recommend platform regulation measures, opposing vague restrictions on information, advocating transparency in content moderation, calling for proportionate enforcement, and emphasizing independent oversight.

Research Questions

RQ1: How do domestic content moderation regulations for combating misinformation differ and align regarding content definition, moderation procedures, transparency obligations, and independent oversight?

RQ2: To what extent do the analysed legal frameworks align with the regulatory recommendations in the UN's 'Countering Disinformation' report and UNESCO's 'Guidelines for the Governance of Digital Platforms' regarding content definition, moderation procedures, transparency, penalties and oversight?

Material and Methods

We conducted a comparative qualitative analysis of domestic jurisdictions regulating content moderation and misinformation. Our approach followed the four-step framework for comparative media policy analysis proposed by Puppis and d'Haenens (2012): case selection, data collection, identification of dimensions, and comparison.

For case selection, we used 'LupaMundi,' a web-based platform developed by the Brazilian fact-checking organisation Agência Lupa (a member of the International Fact-Checking Network), to systematically identify policy responses to online falsehoods. After generating an initial list of legislation, we refined our selection based on four criteria to ensure relevance to our study: (a) domestic laws that target online misinformation and hold platforms accountable, (b) cases within and outside Europe, (c) the influence of these cases on legislation in other countries, and (d) varying degrees of "internet legal content regulation" (University of Gothenburg, 2022). These criteria yielded the previously mentioned five cases. Notably, Germany's NetzDG and France's Law 1202 are early examples of expanding platform liability for content moderation and misinformation, with NetzDG influencing debates beyond Germany (Gorwa, 2024). Although both laws have been partially superseded by the DSA, they remain significant case studies of initial national-level efforts to combat online misinformation. The DSA itself draws on elements from frameworks like NetzDG, highlighting the importance of domestic jurisdictions preceding the DSA's implementation (Haupt, 2024). In this sense, the German and French cases carry important normative weight. Regarding Brazil's resolution, it is the country's only enforceable regulation against online misinformation, following the failure of broader legislative efforts like Bill 2630. Singapore's and Turkey's laws have sparked concern over the expansive powers they grant to state authorities in online content moderation (Ang and Goggin, 2021). Since our criteria did not distinguish between general and election-specific laws, the French and Brazilian cases fall into the latter category (both specifically apply to electoral periods).

For data collection, we retrieved the selected jurisdictions from official sources. Next, the analytical dimensions were deductively established (Tracy, 2013) based on the UN's and UNESCO's documents. Key themes were identified and then grouped into four analytical dimensions: (a) definitions of harmful content and misinformation, (b) content moderation protocols, (c) platform transparency measures, and (d) independent oversight mechanisms. Finally, the comparison involved a software-assisted qualitative data analysis using MAXQDA.

Findings and Discussion

Our findings reveal key differences across the jurisdictions analysed, which can be partially attributed to the regulatory tradition and political context of each country. Countries with the highest scores in "internet legal regulation content" structure their frameworks either around pre-existing legislation, as seen in Germany and France, or around technical criteria that enable forensic verification of misinformation, such as AI-generated content, as observed in Brazil. These countries also impose transparency obligations on platforms regarding content moderation, likely reflecting their freedom of information traditions (for example, the German Freedom of Information Act and Brazil's Law of Access to Public Information). However, their approaches to independent oversight diverge. Germany and France have long-standing traditions of regulating mass media through independent bodies designed to ensure the quality of public communication. Their jurisdictions have extended this model to digital platforms by establishing independent oversight mechanisms to ensure compliance. Conversely, Brazil's media system remains largely unregulated and lacks an oversight body. Similarly, the country's legal framework on online misinformation allows for oversight only through the judiciary, as seen in the Superior Electoral Court's monitoring during election periods. See Table 1 for the core norms.

Table 1[a]: Content moderation norms

Case (Score)[b]	Content definition	Protocols	Transparency	External oversight
Germany (3.73)	Unlawful content according to the German Penal Code. No specific mention to false information (§1-3)	Complaints management system; pre-approved self-regulatory system; removal by platforms (§3)	Report on content removal (§2) + disclosure of data for researchers (§5a)	Regulated self-regulation approach
France (3.9)	False Information. Allegations or inaccurate or misleading imputations (Art. L. 163-2)	Removal by judicial order, upon request from political parties, candidates, etc. (Art. L. 163-2)	Report on measures implemented to combat misinformation and other recommendations (Art. 11)	Arcom
Brazil (3.71)	Manufactured or manipulated content; synthetic content (Art. 9-C)	Removal by platforms and (Art. 9-E) judicial order (Art. 38 §1)	Repository fed by platforms with data on removed content (Art. 9-G §2)	n/a
Singapore (1.54)	False statements of fact (5.)	Correction of false content (11.) by order of government ministers (20.)	n/a	n/a
Turkey (1.66)	Untrue information (Art. 29)	Removal by administrative and court orders (Art. 34)	Report on content deletions and blockades submitted to the authorities (Art. 34)	BTK (independence questioned)

Note: [a]Adapted from Santos, Cazzamatta, and Napolitano (2025, p. 14). [b] Internet legal regulation content. Score 0–4: 0 means the state can remove any content at will; 4 means political speech is protected, andcontent can only be removed if it violates clear, well-established legal standards (University of Gothenburg, 2022).

Countries with lower scores in internet legal regulation content (Singapore and Turkey) also lack independent oversight. In Turkey, although the law grants the BTK (nominally an independent authority) the power to penalise platforms for noncompliance, the agency remains under government control. Both Singapore and Turkey have some of the vaguest legal definitions of misinformation. In Singapore, this broad definition has enabled the government to use the POFMA to issue correction orders against a website that questioned the efficacy of COVID-19 vaccines, as well as against opposition politicians for statements on public policy (Goh et al., 2025). In Turkey, the misinformation law has led to legal consequences for several journalists accused of spreading false information (RSF, 2023). Unlike other regulatory approaches, these laws place

primary enforcement in the hands of government authorities. Consequently, there is less transparency, as platforms are neither mandated nor directly responsible for taking specific actions to manage misinformation.

Conclusion

All jurisdictions analysed struggle, to varying degrees, to balance the removal of online falsehoods with human rights standards. Key concerns include overly vague definitions of false information in certain countries, the absence of judicial or independent oversight in others, and general shortcomings in transparency obligations and user appeal mechanisms. These issues diverge from the recommendations by the UN and UNESCO and raise concerns about potential infringements on free speech. Among the frameworks examined, NetzDG – despite lacking judicial mediation and posing risks of over-blocking due to platform sanctions – aligns more closely with UN and UNESCO standards, particularly regarding transparency and independent oversight. Conversely, Singapore and Turkey diverge further, granting broad powers to state authorities in content moderation with limited safeguards against government overreach. Addressing these gaps requires clearer appeal mechanisms, stronger independent oversight – especially in Brazil, Singapore and Turkey – and greater public involvement in platform accountability. Expanding co-regulatory approaches that truly ensure a balanced distribution of power among multiple stakeholders is a necessary step forward (UNESCO, 2023; Gorwa, 2024; Schuldt, 2021).

Funding

This work was supported by the Deutsche Forschungsgemeinschaft (Project Number 8212383) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

References

Ang, P. H., & Goggin, G. (2021). The regulation of online disinformation in Singapore. In J. Bayer, B. Holznagel, P. Korpisaari, et al. (Eds.), *Perspectives on platform regulation: Concepts and models of social media governance across the globe* (pp. 549–563). Nomos.

Gillespie, T. (2018). Platforms are not intermediaries. *Georgetown Law Technology Review, 2*, 198–216.

Goh, Z. H., Chew, M., & Tandoc, E. (2025). Singapore navigating media regulation and digitization. In A. K. Schapals & C. Pentzold (Eds.), *Media compass: A companion to international media landscapes* (pp. 390–397). Wiley Blackwell.

Gorwa, R. (2024). The politics of platform regulation: How governments shape online content moderation. Oxford University Press.

Haupt, C. E. (2024). Curbing hate speech online: Lessons from the German Network Enforcement Act (NetzDG). In E. Heinze, N. Alkiviadou, T. Herrenberg, et al. (Eds.), *Oxford handbook on hate speech* (Northeastern University School of Law Research Paper No. 467).

Jungherr, A. (2024). Foundational questions for the regulation of digital disinformation. *Journal of Media Law*, *June*, 1–10. https://doi.org/10.1080/17577632.2024.2362484

Puppis, M., & d'Haenens, L. (2012). Comparing media policy and regulation. In F. Esser & T. Hanitzsch (Eds.), *Handbook of comparative communication research* (pp. 221–233). Routledge.

Reporters Without Borders. (2023). Türkiye's year-old "disinformation" law has stepped up pressure on journalists. https://rsf.org/en/türkiye-s-year-old-disinformation-law-has-stepped-pressure-journalists

Santos, A., Cazzamatta, R., & Napolitano, C. J. (2025). Holding platforms accountable in the fight against misinformation: A cross-national analysis of state-established content moderation regulations. *International Communication Gazette, 0*(0), 1–22. https://doi.org/10.1177/17480485251348550

Schuldt, L. (2021). Official truths in a war on fake news: Governmental fact-checking in Malaysia, Singapore, and Thailand. *Journal of Current Southeast Asian Affairs, 40*(2), 340–371.

Tracy, S. J. (2013). Qualitative research methods: Collecting evidence, crafting analysis, communicating impact. Wiley-Blackwell.

UNESCO. (2023). Guidelines for the governance of digital platforms: Safeguarding freedom of expression and access to information through a multi-stakeholder approach. https://unesdoc.unesco.org/ark:/48223/pf0000387339

University of Gothenburg. (2022). *Varieties of democracy (V-Dem) project*. https://v-dem.net

United Nations, Secretary-General. (2022). *Countering disinformation for the promotion and protection of human rights and fundamental freedoms.* https://tinyurl.com/un-disinformation-report

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Navigating the Legal Landscape of APIs

Innovations and Strategies for Sustainable, Inclusive, and Secure Digital Ecosystems

Fabrice Teuguia \\ Cabinet Riding-Up, Programme PCP-ACEFA \\ Yaounde \\ Cameroon
fteuguia@pcpacefa.org

KEYWORDS
API; public interest, legal frameworks, regulatory compliance, fines and penalties, risk mitigation, human well-being

Abstract

APIs (Application Programming Interfaces) have become the backbone of modern software ecosystems, enabling interoperability, innovation, and rapid development. However, the legal landscape surrounding APIs presents complex challenges related to copyright, liability, data protection, and regulatory compliance. This research proposes innovative strategies to create a more secure, ethical, and compliant digital ecosystem through systematic analysis of API legal aspects using TRIZ methodology and scenario planning. The findings provide a roadmap for policymakers, businesses, and developers to maximize API potential while ensuring legal compliance and public interest protection.

Introduction and Problem Statement

The integration of APIs into software development has revolutionized application interactions, enabling seamless data exchange across platforms. Today, APIs power everything from mobile applications to Internet of Things (IoT) devices and cloud services. However, this increased connectivity brings significant legal challenges that organizations must navigate carefully.

The central contradiction identified in this research is the need to reduce risks associated with API usage (security, privacy, compliance) while simultaneously broadening access to APIs for diverse participants in an “Internet of Everything” (universal access, inclusion, and interoperability). This contradiction creates tensions between innovation and regulation, accessibility and security, global standards and local compliance.

Research Methodology

This research employs a multi-faceted approach combining:

- Historical Analysis: Comprehensive review of API evolution and document sharing from legal perspectives (1940s-2024)
- TRIZ Methodology: Application of the Theory of Inventive Problem Solving to identify and resolve contradictions in API legal frameworks
- Laws of System Evolution: Analysis using nine laws of technical system evolution to understand current state and predict future developments
- Scenario Planning and Backcasting: Development of three plausible futures for API governance

The methodology focuses on understanding stakeholder interactions, regulatory evolution, and systemic contradictions within the API legal ecosystem.

Key Findings

System Analysis Using Laws of Evolution

The analysis reveals that public interest acts as the driving force behind API legal frameworks, with regulators and authorities serving as transmission mechanisms. The system demonstrates:

- Complete system structure (1st law respected): All necessary components present
- Energy conductivity (2nd law respected): Efficient energy flow through system parts
- Rhythm harmonization (3rd law not respected): Unsynchronized actions between public interest and regulators
- Increasing ideality (5th law respected): Maximum functionality with minimal negative impacts
- Uneven development (6th law respected): Contradiction between API accessibility and risk reduction

Stakeholder Analysis

Seven key stakeholder groups were identified:

- API developers and owners: Ensuring compliance and clear usage agreements
- API consumers: Adhering to terms, respecting IP rights, ensuring data security
- Regulators and authorities: Overseeing compliance with data protection and industry standards
- End users/IP owners: Protecting personal data rights and privacy
- Legal professionals: Navigating complex legal landscapes
- Business stakeholders: Safeguarding operations and intellectual property
- Third-party service providers: Building trust and ensuring regulatory compliance

Innovation Propositions

Based on TRIZ analysis, five key innovations are proposed:

Redefining Public Interest Parameters

Shift from traditional regulatory frameworks to inclusive, adaptive models that:

- Implement RegTech solutions for real-time compliance monitoring
- Ensure transparent API usage with clear data practice disclosures
- Promote sustainable API use through public-private partnerships
- Define clear legal aspects of API monetization to prevent exploitation

Segmenting Public Interest Considerations

Break down public interest into manageable segments:

- Stakeholder Groups: Developers, end-users, businesses, marginalized communities
- Geographic Regions: Local, national, international levels
- Sector-Specific Needs: Healthcare, education, finance
- Emerging Trends: AI ethics, environmental sustainability, digital accessibility

Dynamic Organizational Transformation

Replace mechanical, static regulatory organizations with dynamic, specialized systems that:

- Continuously reassess public interest parameters
- Create specialized regulatory panels for different API aspects
- Implement real-time monitoring using AI and machine learning
- Adopt flexible enforcement mechanisms adapted to specific circumstances

Preliminary Implementation Actions

Four foundational steps for modernizing API legal landscapes:

- Audit existing laws and regulatory practices
- Form diverse working groups with multiple stakeholder perspectives
- Build technical expertise within regulatory bodies
- Educate the public about API importance and legal aspects

Specialized Enforcement Mechanisms

Shift responsibility for issuing fines from regulatory bodies to specialized enforcement agencies, allowing regulators to focus on policy-making, oversight, and stakeholder engagement.

Scenario Analysis

Three plausible future scenarios were developed:

Harmonized Ecosystem Scenario (Optimistic)

Characteristics: Global API standards, aligned legal frameworks, strong public-private partnerships Outcomes: High interoperability, moderate regulatory oversight, balanced innovation, high security, moderate compliance costs Key Actions: Establish global standards, foster multi-stakeholder collaboration, implement adaptive regulatory mechanisms

Fragmented Regulation Scenario

Characteristics: Localized compliance, disconnected frameworks, complex compliance landscape Outcomes: Low interoperability, high regulatory oversight, limited innovation, uneven security, high compliance costs Key Actions: Develop localization strategies, promote regional collaboration, establish dynamic compliance mechanisms

Open Innovation Ecosystem Scenario

Characteristics: Collaborative governance, open standards, inclusive innovation, dynamic regulation Outcomes: Moderate interoperability, low regulatory oversight, high innovation, low-moderate security, low compliance costs Key Actions: Establish universal standards, promote collaborative governance, foster inclusive innovation

Practical Implications and Recommendations

For Policymakers

- Implement adaptive regulatory frameworks that can respond quickly to technological changes
- Foster international cooperation for harmonized API governance
- Establish specialized agencies for API compliance enforcement
- Create sandboxes for testing innovative API solutions

For Businesses

- Invest in RegTech solutions for automated compliance monitoring
- Develop clear API governance frameworks addressing all stakeholder needs
- Engage proactively with regulatory bodies and public interest groups
- Implement privacy-by-design principles in API development

For Developers

- Embed standardized security and privacy features in API design
- Maintain transparent documentation of API data practices
- Ensure accessibility and inclusivity in API development
- Stay updated on evolving legal requirements across jurisdictions

Conclusion and Future Research

This research demonstrates that the legal aspects of APIs require ongoing attention and adaptation to address emerging challenges. The proposed innovations provide a comprehensive framework for creating secure, ethical, and compliant digital ecosystems that balance innovation with public interest protection.

The analysis reveals that current regulatory systems often lag behind technological advancement, creating governance gaps. By implementing the five proposed innovations—redefining public interest parameters, segmenting considerations, transforming organizational structures, taking preliminary actions, and specializing enforcement—stakeholders can create more effective API governance frameworks.

Future research should focus on:

- Empirical testing of proposed regulatory innovations
- Development of automated compliance tools using AI
- Cross-cultural studies of public interest definitions in API contexts
- Long-term impact assessment of different scenario implementations

The journey toward effective API governance requires collaboration among all stakeholders, continuous adaptation to technological changes, and unwavering commitment to public interest while fostering innovation. As we advance into 2025 and beyond, these integrated strategies will be essential for building resilient ecosystems that support both technological advancement and societal well-being.

Selected References

Digital Government of New Zealand. (2022). *API guidelines — Part A: API concepts and management 2022*. https://www.digital.govt.nz

Fielding, R. (2000). *Architectural styles and the design of network-based software architectures* [Doctoral dissertation, University of California, Irvine]. https://www.ics.uci.edu/~fielding/pubs/dissertation/fielding_dissertation.pdf

Koch, R. (2024). Big Tech has already made enough money in 2024 to pay all its 2023 fines. *Proton*. https://proton.me/blog/big-tech-fines-2024

McCormack, M. (2024). The importance of regulatory compliance risk assessments. *Compliancy Group*. https://compliancy-group.com

Sagdeo, P. (2018). Application programming interfaces and the standardization-value appropriation problem. *Harvard Journal of Law & Technology, 32*(1). https://jolt.law.harvard.edu/assets/articlePDFs/v32/Sagdeo-Application-Programming-Interfaces.pdf

Thadchanamoorthy, V. (2024). APIs and data protection. *ActiveMind Legal*. https://www.activemind.legal

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Politicising Digital Policy in the European Union

Ben Crum \\ Vrije Universiteit Amsterdam \\ Amsterdam \\ Netherlands
ben.crum@vu.nl

KEYWORDS
EU digital policy; politicisation; depoliticisation; digital democracy; digital sovereignty

Introduction

One can hardly say that digital policy is apolitical these days. Still, the terms in which we understand the politics of digital technology have become heavily skewed towards geo-politics. Much of the debate is framed by the technological competition between the United States and China, among which Europe has to fight for its own position or become dependent on them (Seidl & Schmitz 2023). This focus on the *geo*-politics of digital policy risks suppressing the political issues and divisions that digital technology raises *within* the EU.

This paper draws on democratic theory, analyses of on (de-)politicisation in the EU, and instances of recent EU digital legislation to identify the kind of mechanisms that (may) prevent the politicisation of digital policy in the EU. Specifically, I want to highlight how, while some of these mechanisms have been well-established for the EU in general, they are being reinforced by features that are specific to the domain of digital policy.

Theory: Mechanisms of Depoliticising Digital Policy

The political debate on digital technology raises many issues that transcend the scope of existing political cleavages. However, the current political divisions in the EU's legislative institutions do not always allow these concerns to be articulated in full. It seems even less likely that the terms of these debates reach the EU publics at large.

Some of the reasons why EU digital policy-making remains under-politicised apply to EU policy-making at large (Schmidt, 2006; Mair, 2013: chap. 4; Scharpf 2015). However, these depoliticisation

mechanisms are complemented and amplified by specific features of the digital domain. In this paper, I highlight three such mechanisms.

The first is the complexity and uncertainty surrounding much of the technology. Lack of understanding the underlying technology prevents many members of the public from taking a stance on the issue. It also creates considerable space for those who can claim some authority (above all the tech firms involved) to dominate the debate (Bouza García & Oleart, 2024). The sense of detachment is reinforced by the fact that most of the public have come to see digital technology primarily as a private product, where their autonomy is mostly expressed through contracting in or out, rather than as a public infrastructure on which they might lay a democratic claim ('exit' over 'voice' in Hirshman's (1970) classical terminology).

The second specific feature that hinders the politicization of EU digital policy is that it has come to be dominated by an 'us-versus-them' logic. This framing initially very much took the form of standing up for EU values against the commercial and exploitative interests of Big Tech. Over time, it also has taken on a more geopolitical orientation in which the EU is supposed to assert its (digital) sovereignty against its main economic rivals, the US and China (Pohle & Thiel, 2020; Seidl & Schmitz, 2023). Obviously, these antagonists tend to coincide as most of the Big Tech companies come from the US and, to a lesser extent, China, and the desire of the EU to disentangle from them has become more urgent since the start of the second presidency of Donald Trump.

The other side of this logic is that digital technology and the values that this technology is meant to serve are often presented as uncontested 'European public goods' (Bria, Timmers & Gernone, 2025). This tendency has been particularly reinforced by the Draghi-report (Draghi, 2024). However, such approaches downplay the choices involved in charting the EU's digital domain and the competing interests that different societal strata may have in it. A shared and unified European interest in digital technology cannot be presumed; even if it can be created, it can only come around through democratic deliberation.

A final challenge that is inherent in EU digital policy is that it is arguably the first policy domain in which member states have had no preceding national policies before they had to engage with EU-level decision-making. In a way, this is an advantage as member states have little entrenched stakes to protect. But it also means that they cannot depart from a starting position that has already been legitimated by a process of democratic decision-making at home and by well-trusted practices and expectations.

This claim may not (fully) apply to the largest member states, like Germany (Gorwa, 2024: chap. 5) and France (Bigot, 2022). However, for most EU member states it is more or less self-evident that the main strategic orientations on digital policy are set at the EU level. Recognizing that digital markets know no boundaries and much of the technology originates from the US, the average member state lacks the means to impose its own distinct political preferences. What is more, as digital services are very much part of the single market, they naturally fall under the jurisdiction of the EU.

As a consequence, we see the construction of a body of European law and policies on digital technology that is short of anchoring in national debates and public spheres. This is not only due to the

well-established dynamics of the under-politicization of EU decision-making and the two-level games that undermine full engagement of national democratic institutions, but also because there are no articulated national positions to begin with. What is more, the depoliticization of EU digital policy is further amplified because of the technological complexity of the matter involved and because of the pressure to maintain a united European front against foreign domination through digital technology.

Empirical Probe: The DSA/DMA Package in the National Press

By way of a first probe of the preceding argument, I analyse the coverage of the decision-making on the DSA/DMA-package in three major broadsheets in Germany, France, and the Netherlands (Table 1). While there is little doubt that most EU citizens remain little informed about the EU's decisions in the domain of digital policy and the political choices made in that regard, I find rather distinct patterns between the three countries. Typically, the German newspaper gives more attention to the competition-oriented DMA, while attention for the DSA is bigger in the French and the Dutch newspaper. More importantly, I find evidence that mentions of the EU acts go up when they can be connected to national policy-making or to national politicians, like Thierry Breton and the Council presidency in the French case. This effect seems more likely for bigger member states and for states that have a developed regulatory apparatus for the digital sphere of their own.

Table 1: Articles on DSA/DMA in three European newspapers (1 Dec 2020 – 1 Dec 2022)

	Total articles	Politically relevant*	Visible conflict/ choice
Süddeutsche Zeitung	53	36 (68%)	12 (23%)
Le Figaro	114	58 (51%)	22 (19%)
NRC Handelsblad	31	18 (58%)	7 (23%)
All the newspapers	198	112 (57%)	41 (21%)

Note: * 'Politically relevant' are articles that identify specific EU actors; it excludes articles that only mention "the EU" in general and/or and mostly report on digital firms or national policy.

The findings also show that the Commission is the most covered EU actor, but that its decisions are hardly, if at all, politicised. The newspaper reader also gets little insight in the politics among the national governments in the Council. Only *Le Figaro* offers a spatter of hints. In the end, it is the coverage of the debates in the European Parliament that gives most direct insight in some of the main political choices at stake: targeted advertising, interoperability, and the scope of application of the gatekeeper provisions.

Concluding Comments

These empirical findings offer far from conclusive proof of my theoretical argument that the politicisation of EU digital policy faces particular challenges on top of the usual hurdles that apply in EU politics. However, some of the evidence lends support to the argument that it is more difficult to politicise an EU issue in the absence of a mature national policy on the issue. Also, some of the findings seem to support the thesis that digital issues are more likely to be framed as pitting a (presumed) collective EU interest against (foreign) Big Tech than as involving trade-offs among competing interests among EU citizens. These are relevant hurdles to consider if EU publics are to have a genuinely democratic discussion about how the digital domain is to be regulated.

References

Bigot, C. (2022). La liberté de communication dans la loi du 24 août 2021, les nouvelles obligations de collaboration des plateformes sous le contrôle de l'ARCOM. *Légipresse, 67*(HS3), 31–43. https://doi.org/10.3917/legip.hs67.0031

Bouza García, L., & Oleart, A. (2024). Regulating disinformation and Big Tech in the EU: A research agenda on the institutional strategies, public spheres and analytical challenges. *JCMS: Journal of Common Market Studies, 62*, 1395–1407. https://doi.org/10.1111/jcms.13548

Bria, F., Timmers, P., & Gernone, F. (2025). *EuroStack – A European alternative for digital sovereignty.* Bertelsmann Stiftung. https://www.euro-stack.info

Draghi, M. (2024). *The future of European competitiveness.* European Commission. https://commission.europa.eu/topics/eu-competitiveness/draghi-report_en

Gorwa, R. (2024). *The politics of platform regulation: How governments shape online content moderation*. Oxford University Press.

Hirschman, A. (1970). *Exit, voice, and loyalty.* Harvard University Press.

Mair, P. (2013). *Ruling the void: The hollowing of Western democracy*. Verso Books.

Pohle, J., & Thiel, T. (2020). Digital sovereignty. *Internet Policy Review, 9*(4), 1–19. https://doi.org/10.14763/2020.4.1532

Scharpf, F. (2015). Multilevel European democracy. *European Law Journal, 21*(3), 384–405. https://doi.org/10.1111/eulj.12127

Schmidt, V. (2006). Democracy in Europe: The EU and national polities. Oxford University Press.

Seidl, T., & Schmitz, L. (2023). Moving on to not fall behind? Technological sovereignty and the 'geo-dirigiste' turn in EU industrial policy. *Journal of European Public Policy, 31*(8), 2147–2174. https://doi.org/10.1080/13501763.2023.2248204

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Regressive ‚Alternative' Media and Their Role in Disrupting the Public Spheres in Europe and Latin America

Regina Cazzamatta \\ Universität Erfurt \\ Erfurt\\ Germany
regina.cazzamatta@uni-erfurt.de

KEYWORDS
fact-checking; mis- and disinformation; digital native outlets; regressive 'alternative' websites; changing meaning of 'alternative'; comparative studies

Abstract

The emergence of alternative media traces back to the progressive social movements of the 1960s and 1970s, particularly on the progressive left, which sought to democratize media production and amplify underrepresented perspectives (Blöbaum, 2006; Harlow, 2023; Klawier et al., 2023). Recently, new online platforms have emerged, self-proclaiming as "alternative" media and challenging the establishment but often influenced by far-right ideologies. In Latin America, some "digital natives" view themselves as independent counterweights that address public issues overlooked by traditional media but vehemently rejecting the "alternative" label (Harlow, 2023).

In media and communication scholarship, when analyzing the networked public sphere and hybrid media systems, scholars (Chadwick et al., 2017; Kaiser et al., 2018) often use the term 'alternative' in reference to mainstream news institutions without critically examining power dynamics. According to Jackson and Kreiss (2023), labeling the far-right as a counter-public under the banner of 'alternative' can equate the most undemocratic groups with those aiming to enhance democracy. The global right-wing movement should be more accurately termed "defensive" publics due to their alignment with systemic inequalities. Unlike progressive social movements, far-right groups seek to preserve "exclusionary racial, social, political, and economic orders" (Jackson & Kreiss, 2023, p. 103).

We contend that it is not just the right wing's alignment with systemic inequalities but also their dubious relationship with truth and facts that set them apart from the conventional notion of 'alternative' publics. While independent digital native projects often integrate fact-checking units to

combat disinformation (Cazzamatta, 2025b; Graves, 2018; Lilienthal, 2017), right-wing outlets often serve as significant sources of falsehoods. These websites not only employ democratic rhetoric but also undermine constructivist perspectives on facts in their pursuit of anti-democratic goals.

In his essay titled "Why Has Critique Run out of Steam? From Matter of Facts to Matter of Concern," French sociologist Bruno Latour questions what has gone awry, particularly in light of climate change denialists' arguments grounded in social constructivist views (Latour, 2004).

This article, recently published in *Publizistik* (Cazzamatta, 2025a), examines the mission statements of regressive alternative media identified by fact-checkers in 2022 across eight countries in Europe and Latin America: Germany, the UK, Portugal, Spain, Brazil, Argentina, Chile, and Venezuela. These countries have varied media landscapes, political systems, degrees of political polarization, approaches to populist communication, and patterns of social media use for news. Considering the international scope of these outlets, the study initially mapped the primary locations of misinformation sources identified by fact-checkers in these regions and analyzed how these websites justify their societal roles through their mission statements.

Identified through extensive content analysis of fact-checking articles under a DFG project titled "[Disinformation Landscapes and the Emergence of Fact-Checking Organizations in Europe and Latin America]," our methodology involves qualitatively analyzing these websites' mission statements and slogans to uncover the strategies used to legitimize their activities. Among the primary findings, regressive alternative media are more established in Germany and Spain, where far-right movements are more prominent online (Cazzamatta, 2024). In Latin America, while such outlets are less widespread, the continent nevertheless experiences significant impact from regressive websites based primarily in Spain and the US, highlighting their global reach in spreading misinformation (Heft et al., 2021).

It is also plausible that radical right-wing discourses are less marginalized in Latin American public debates due to the conservative stance of legacy media and the reciprocal relationship between the state and media (Harlow, 2023; Waisbord, 2000), which limits the growth of regressive niche media (Heft et al., 2020). Despite varying missions, these European and, to a lesser extent, Latin American outlets generally promote exclusionary and regressive views, advocating for a more traditional, less inclusive social order where freedom of expression is reserved for select groups (della Porta, 2023). Although fact-checkers have identified progressive outlets spreading misinformation, these cases are rare, with only two found—one in Germany and one in Spain.

References

Blöbaum, B. (2006). Wandel alternativer Öffentlichkeit: Eine Fallstudie zur *tageszeitung (taz)*. In K. Imhof, R. Blum, H. Bonfadelli, & O. Jarren (Eds.), *Demokratie in der Mediengesellschaft* (pp. 182–192). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-90511-2_11

Cazzamatta, R. (2024). Global misinformation trends: Commonalities and differences of topics, sources of falsehoods and deception strategies across eight countries. *New Media & Society, 0*(0). https://doi.org/10.1177/14614448241268896

Cazzamatta, R. (2025a). Regressive 'alternative' media and their role in disrupting the public spheres in Europe and Latin America. *Publizistik, 1/2*. https://doi.org/10.1007/s11616-025-00892-9

Cazzamatta, R. (2025b). The content homogenization of fact-checking through platform partnerships: A comparison between eight countries. *Journalism & Mass Communication Quarterly, 102*(1), 120–157. https://doi.org/10.1177/10776990241261725

Chadwick, A., Dennis, J., & Smith, A. P. (2017). Politics in the age of hybrid media: Power, systems, and media logics. In A. Bruns, G. Enli, E. Skogerbø, A. O. Larsson, & C. Christensen (Eds.), *Routledge companion to social media and politics* (pp. 7–22). Routledge.

della Porta, D. (2023). Regressive movements in times of emergency: An introduction. In D. della Porta (Ed.), *Regressive movements in times of emergency* (1st ed., pp. 1–37). Oxford University Press. https://doi.org/10.1093/oso/9780198884309.003.0001

Graves, L. (2018). Boundaries not drawn: Mapping the institutional roots of the global fact-checking movement. *Journalism Studies, 19*(5), 613–631. https://doi.org/10.1080/1461670X.2016.1196602

Harlow, S. (2023). Digital-native news and the remaking of Latin American mainstream and alternative journalism. Routledge.

Heft, A., Knüpfer, C., Reinhardt, S., & Mayerhöffer, E. (2021). Toward a transnational information ecology on the right? Hyperlink networking among right-wing digital news sites in Europe and the United States. *The International Journal of Press/Politics, 26*(2), 484–504. https://doi.org/10.1177/1940161220963670

Heft, A., Mayerhöffer, E., Reinhardt, S., & Knüpfer, C. (2020). Beyond Breitbart: Comparing right-wing digital news infrastructures in six Western democracies. *Policy & Internet, 12*(1), 20–45. https://doi.org/10.1002/poi3.219

Jackson, S. J., & Kreiss, D. (2023). Recentering power: Conceptualizing counterpublics and defensive publics. *Communication Theory, 33*(2–3), 102–111. https://doi.org/10.1093/ct/qtad004

Kaiser, J., Fähnrich, B., Rhomberg, M., & Filzmaier, P. (2018). What happened to the public sphere? The networked public sphere and public opinion formation. In E. G. Carayannis, D. F. J. Campbell, & M. P. Efthymiopoulos (Eds.), *Handbook of cyber-development, cyber-democracy, and cyber-defense* (pp. 433–459). Springer International Publishing. https://doi.org/10.1007/978-3-319-09069-6_31

Klawier, T., Prochazka, F., & Schweiger, W. (2023). Public knowledge of alternative media in times of algorithmically personalized news. *New Media & Society, 25*(7), 1648–1667. https://doi.org/10.1177/14614448211021071

Latour, B. (2004). Why has critique run out of steam? From matters of fact to matters of concern. *Critical Inquiry, 30*(2), 225–248. https://doi.org/10.1086/421123

Lilienthal, V. (2017). Recherchejournalismus für das Gemeinwohl: Correctiv – eine Journalismusorganisation neuen Typs in der Entwicklung. *Medien & Kommunikationswissenschaft, 65*(4), 659–681. https://doi.org/10.5771/1615-634X-2017-4-659

Waisbord, S. (2000). Media in South America: Between the rock of the state and the hard place of the market. In M.-J. Park & J. Curran (Eds.), *De-westernizing media studies* (pp. 43–53). Routledge.

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Platform-Government Power Dynamics and the Governance of Disinformation

Regulatory and Policy Frameworks in Egypt

Maysa Amer \\ Freie Universität Berlin \\ Berlin \\ Germany
maysa.amer@fu-berlin.de

KEYWORDS
platform, governance; disinformation; power; platform, regulations; global, south; misinformation; internet, governance; media, policy; social, media

Abstract

Disinformation and misinformation are widespread global issues that affect all countries and adversely impact several facets. Disinformation has been immensely researched across the globe. However, there is a scarcity of research that has expanded its scope to encompass countries in the global south and the Middle East, a region characterized by extensive use of digital platforms like social media and a lack of measures to safeguard citizens against deceptive online activities. This study contributes to the broader literature that examines disinformation in authoritarian regimes and the global south, its influences, and countermeasures by examining the relevant platform regulations and media policies.

Building on power theories (Lukes, 2005, 2021; Gaventa, 2006), this study aims to gain an in-depth understanding of the power dynamics in which disinformation emerges, disseminates, proliferates, and is regulated. By illustrating the Middle East region, with an emphasis on Egypt, and considering the geopolitical and sociopolitical aspects, the study looks into how disinformation actors and power dynamics among stakeholders influence content moderation practices on social media platforms. In addition, the study examines platform regulations and media policies in place addressing disinformation and misinformation and investigates their implications.

Thus, this project seeks to analyze 1) the current regulatory framework of disinformation in Egypt and the mechanisms by which the governments intervene in content moderation and content removal, 2) the power dynamics among stakeholders, including governments and platforms over content moderation practices, and 3) the extent to which disinformation and its countermeasures have

ramifications and influence trust towards digital content. In addition to investigating digital policies and laws that address disinformation, the study looks into how governments/states could intervene and play a key role in platform governance from a national standpoint. The study employs a triangulation of qualitative research methods, which include text analysis for laws and regulations in relevance, document analysis, secondary resources, and 32 in-depth interviews with officials, policymakers, experts, policy representatives from social media platforms, and media professionals.

The study aims to answer three main research questions as follows:

- What are the current platform regulations and media policies in Egypt addressing disinformation and misinformation?
- How to explain the role of stakeholders in regulating disinformation and the dynamics of their power?
- What are the patterns of disinformation, their countermeasures, and their implications?

The study contributes to the global scientific body of research on platform governance in general and governance of disinformation and misinformation in particular by examining the dissemination and governance of disinformation from a national perspective. The hard-to-access data obtained in this study enriches our comprehension of the actors involved in disinformation, the regulatory frameworks governing it, and the ramifications for free speech, especially in light of the ongoing usage of AI tools that proliferate the dissemination of disinformation even further.

This abstract was accepted for the conference but has not been presented in person.

Funding

This project received funding from DAAD, the German Academic Exchange Service.

References

Gaventa, J. (2006). Finding the spaces for change: A power analysis. *IDS Bulletin, 37*(6), 23–33. https://doi.org/10.1111/j.1759-5436.2006.tb00320.x

Lukes, S. (2005). Power and the battle for hearts and minds. *Millennium, 33*(3), 477–493. https://doi.org/10.1177/03058298050330031001

Lukes, S. (2021). Power and domination. In *Essays on evolutions in the study of political power* (pp. 97–108). Routledge.

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Participatory Propaganda and the Affordances of Digital Platforms

Explaining the Emergence of Far-Right Islamist Protest Mobilization in Indonesia

Salma Aqida \\ Goethe University Frankfurt \\ Frankfurt am Main \\ Germany
aqidasalma@stud.uni-frankfurt.de

KEYWORDS
participatory propaganda; digital media; social movements; far-right activism; islamism

Abstract

The 2016 far-right Islamist demonstrations in Jakarta, known as the "Action to Defend Islam," are largely regarded as the largest in Indonesian history since the fall of Suharto in 1998. Despite being historically marginalized and operating largely outside mainstream politics, the far-right Islamist movement was able to mobilize an extraordinarily large crowd. While some argue that the demonstrations indicate a growing anti-pluralist and illiberal sentiment among Indonesian Muslims, others credit social media for enabling the successful mobilization, comparable to its role in the Arab Spring. In this article, I argue that these demonstrations are best understood through the lens of participatory propaganda, facilitated by digital platforms. Departing from prior studies focusing on the Western world and specific political actors, this research examines the potential connection of participatory propaganda to protest movement mobilization. Identifying six digital tactics employed by far-right actors and supporters, I illustrate how they seed and shape participatory propaganda cascades, ultimately leading to and sustaining their mobilization. I posit that the unprecedented mobilization during the Jakarta election is not solely due to growing conservatism among Indonesians but is rather a result of the new information environment that has drastically altered the ways individuals engage with and are influenced by public information. By examining both social media content and delivery, the study reveals how digital media affordances enable varying degrees of awareness for propaganda, turning individuals into both producers and spreaders, ultimately unveiling a hybrid form of propaganda that illustrates the interplay of new and traditional media, as well as online and offline activism.

Introduction

During the 2016 Jakarta election, Indonesia witnessed a remarkable mobilization driven by radical right and conservative Islamist groups, notably the Islamic Defenders Front (FPI), known as the Action to Defend Islam (or in Bahasa Indonesia called Aksi Bela Islam abbreviated ABI). This mobilization aimed to express Islamist opposition to Basuki Tjahaja Purnama, known popularly as Ahok, a double minority Chinese-Christian former governor of Jakarta, the Muslim-majority capital city of Indonesia. The largest demonstration, which took place in December 2016, attracted an unprecedented 800,000 protestors, making it the most significant protest in the country since the 1998 demonstrations that helped topple the dictatorial Suharto regime (Lindsey, 2016). The protestors demanded the arrest of Ahok for allegedly violating the Quran. While the movement claimed to be "defending Islam," it also directed animosity towards individuals, religions, and ethnicities (Hadiz, 2019). This dual nature attracted significant attention among analysts, particularly for the religious polarization it generated (e.g., Lim 2017). Despite historically being marginalized and operating outside mainstream politics, the far-right Islamist movement was able to mobilize to an extraordinary extent. Even government and security officials were taken aback by the sheer magnitude of the crowd and the seriousness of their demands during these protests.

The mobilization of ABI began with a heavily edited video of the incumbent Jakarta governor on Facebook, which included text transcripts and captions suggesting that Ahok purposefully disrespected Islam. On September 27, 2016, while giving a speech in the Kepulauan Seribu (Thousand Islands), Ahok criticized his political opponents for using Islam as a campaign tool. He stated that voters were being “deceived using Verse 51 of al-Maida” (a chapter of the Quran)1. After being liked and shared by netizens, the video quickly went viral, inciting outrage among the masses and influential Muslim figures. Ahok’s statement was utilized by Islamist groups to mobilize anger among Muslims by framing it as an insult of Islam. Despite enjoying high approval ratings as the incumbent, Ahok was defeated in the gubernatorial race. Subsequently, he was sentenced to a two-year prison term for blasphemy, with the edited video of his speech serving as incriminating evidence.

Many credit digital media for enabling the successful ABI mobilization, comparable to its role in the Arab Spring (Howard & Hussain, 2013). These events have prompted scholars to explore the role of digital platforms as essential elements in understanding protest movement mobilization. This examination is grounded in the emerging logic of aggregation, as proposed by Juris (2012), which underscores the importance of digital platforms and social media in bringing together diverse actors to form cohesive social movements or collective actions. Additionally, scholars have coined terms such as 'connective action' (Bennett & Segerberg, 2012), 'digitally networked action' (Bennett & Segerberg, 2011), 'e-movements' (Earl & Kimport, 2011), or 'cloud protesting' (Milan, 2011) to describe digitally enabled collective action and social movements. This phenomenon extends to far-right

1 The verse to which Ahok referred advises against aligning with Christians and Jews. However, Ahok did not explicitly name the individuals he accused of using Quranic verses to dissuade people from voting for him. To prevent misinterpretations of sacred texts, adherents of religious faiths should endeavor to understand the contexts in which these texts originally appeared.

movements, which often encounter challenges in gaining coverage or maintaining a presence in mainstream mass media channels (Jost & Dogruel, 2023), where gatekeeping can impede the spread of their radical ideologies (Kaiser et al., 2019). As a result, they increasingly rely on digital platforms to overcome these mobilization constraints in the public sphere. Social media serves as a vital tool for them to disseminate their ideas and engage with the public quickly and cost-effectively (Caiani & Kröll, 2014;), a motivation that ABI activists also recognized for their strategic use of these platforms.

One relevant example is the riot at America's Capitol on January 6th, 2021, which was at least partly born on social media (Kydd, 2021). Nowadays, it is difficult to envision figures like Trump without Twitter, Bolsonaro without YouTube, or, within the context of this research, the ABI movement without Facebook. In Indonesia, since the early days of the Internet in the 2000s, with chat rooms, mailing lists, and websites, far-right Islamists have adeptly exploited technological advances. This is evident in the case of Laskar Jihad Online (Bräuchler, 2004; Lim, 2005), which sought to proliferate Islamic (jihad) websites, with http://www.laskarjihad.or.id serving as the official site, alongside Yahoo-based mailing lists boasting thousands of members spread across Indonesia. The 2016 demonstrations were thus exemplary of a more general entanglement between the rise of far-right Islamist mobilization and digital technology in contemporary Indonesia.

However, it remains unclear how one, rather clearly and heavily manipulated video led such a large number of people to participate in a protest organized by extremist and fringe Islamist groups that were not previously aligned with the mainstream ideologies prevalent among Indonesia's Muslim-majority population. The question arises: How did these far-right Islamist organizations manage to mobilize such a vast and diverse group of Indonesians? Does the rise of the far-right Islamist movement truly align with the growing conservatism of Indonesian Muslims? What explains the success of this mobilization? What are its causes and what can we expect in the future? Furthermore, how did digital technology contribute to facilitating this mobilization? Investigating the trajectory of this specific movement is essential because it defies conventional expectations of far-right Islamist movements and indicates a shift in the dynamics of Islamist movements in Indonesia and their ability to mobilize large numbers of supporters through social and political means.

In this article, I argue that the successful ABI mobilization can be explained by combining the concept of participatory propaganda (Jowett & O'Donnell, 2015; Wanless & Berk, 2017; Lewandowsky, 2022) with the specific affordances of digital platforms. Propaganda is traditionally defined as the dissemination of persuasive information in order to influence a target audience to adopt behaviors or beliefs favored by the propagandist (Bernays, 1928; Lasswell, 1948). In the digital age, modern propaganda has evolved to become increasingly participatory, involving collaboration with online crowds rather than simply disseminating messages in a one-way manner. This participatory nature transforms the targets of propaganda into active collaborators, significantly altering the dynamics of information dissemination. This shift is particularly notable in the context of far-right groups, as recent research underscores the significant role of user engagement in right-wing radicalization processes (Calvert, 2024). User activity plays a pivotal role in determining the speed and extent to which radical ideology and propaganda can spread across platforms, highlighting the importance of understanding and leveraging participatory dynamics in contemporary propaganda efforts.

This collaborative element is distinct from the participatory propaganda practices of the early 2000s Internet age in Indonesia. While the propaganda of Laskar Jihad Online was also participatory in nature, the emergence of social networking platforms has introduced new dynamics. During the Laskar Jihad mobilization, citizens were primarily encouraged to produce propaganda content through non-digital autonomous media such as radio broadcasts, magazines, books, newsletters, VCDs, pamphlets, and print newspapers, among other formats. In contrast, digital co-production today is less hierarchical, allowing netizens greater freedom to express their opinions and creative ideas as part of the persuasion project. Additionally, the speed of social media content production facilitates the interactive service dimension of participatory persuasion, a practice less feasible in the early Internet Laskar Jihad era. This interactive dimension involves real-time interactions on social media platforms, where users engage with propaganda content through actions like liking, sharing, commenting, and creating their own content. Such engagement fosters active participation in spreading propaganda messages and enhances the effectiveness of participatory persuasion. Unlike in the past, we are now seeing a more subtle form of participatory persuasion, where netizens are subtly influenced without always being fully aware of it.

The concept of participatory propaganda echoes some scholarship on China's authoritarian regime, where it is termed "authoritarian participatory persuasion 2.0," highlighting the shift towards a more participatory form of digital persuasion intended to facilitate public complicity with the regime under President Xi (Repnikova & Fang, 2018). Similarly, studies in the Western context explore the participatory propaganda model, with a focus on Western political actors and campaigns. For instance, Wanless and Berk (2020) examined the rise of participatory propaganda during the 2016 US presidential election, a trend that was later observed in online political activities in the UK and Canada. Specifically, Starbird (2023) studied how propaganda materials were co-created and disseminated on Twitter by political elites, media figures, activists, and online users, contributing to the spread of misinformation about voter fraud that ultimately led to violent events at the US Capitol.

This article extends this literature by examining the connection between participatory propaganda and protest movement mobilization in the Global South, demonstrating the global nature of this phenomenon. It thus provides further evidence for the emerging view that in the 21st century, propaganda is evolving from persuasion to participation, becoming a more potent tool for shaping collective behavior. This perspective encourages us to broaden our focus beyond computational propaganda, which typically focuses on actors like "bots," "buzzers," and "trolls" (Woolley & Howard, 2016; Sastramidjaja & Wijayanto, 2022; Rasidi, 2023; Ledford, 2024). Instead, we should consider more complex conceptualizations that account for how propaganda integrates into online communities and audiences, whether they are aware or unaware, as both recipients and contributors to propagandistic content generation and dissemination. In this work, I demonstrate how far-right Islamist groups leverage the participatory nature of digital communication by fostering collaborative efforts among various groups, including activists, political elites, hyper-partisan media, social media influencers, and ordinary citizens to achieve mobilization goals. This approach allows for a more nuanced understanding of how propaganda operates, including both the intentional efforts of paid

actors and the unintentional coordination among individuals who may not share the same goals or even be aware of their role in the mobilization.

In what follows, I will identify six digital tactics employed by far-right strategic actors and their supporters. These tactics are instrumental in initiating and shaping participatory propaganda cascades, which ultimately drive multiple mobilization successes. Through the systemic integration of various techniques facilitated by technological advancements, the strategic use of behavioral sciences within a participatory propaganda model, and the emerging logic of aggregation in social movements, propagandists can deliberately target specific audiences.

Beyond Manipulation: How Digital Platforms Reshaped Propaganda into a Participatory Process

Propaganda, once viewed as a top-down, centralized effort to manipulate the public, has evolved significantly in the digital age. While classical theorists like Lasswell and Bernays saw propaganda as a strategic tool used by states and large institutions, this model is no longer sufficient. Today, propaganda is a more fluid and decentralized form of communication. Traditional distinctions between a "propagandist" and an "audience" are collapsing as digital platforms enable individuals to become active participants. Modern propaganda invites engagement, thriving on user-generated content, and spreading through interpersonal networks and algorithmic amplification. This participatory nature means propaganda is often co-created, with its influence growing as users actively shape, repeat, and reinforce the message.

The shift to participatory propaganda is particularly evident in movements like QAnon and Indonesia's Aksi Bela Islam (ABI). These movements demonstrate that ideological momentum doesn't require a strong, centralized authority. Instead, it can build from fragmented and loosely coordinated networks. Propaganda in this context isn't about converting audiences to a single, coherent ideology but rather about activating existing ideological positions and emotions. It operates by reframing events to fit pre-existing worldviews, using repetition and emotional resonance to transform beliefs into entrenched opinions and moral imperatives. For example, the claim that the 2020 U.S. election was 'stolen' or that Ahok 'insulted Islam' becomes not just a statement of belief but a moral imperative tied to identity, belonging, and political action. Once internalized, these narratives justify collective action as a rational and even necessary response. This process turns users from passive recipients into active propagators, making them unwitting collaborators in a broader ideological project.

This evolution is closely tied to the logic of aggregation, a concept describing how digital platforms allow dispersed individuals to rapidly form large-scale collective entities. This logic prioritizes flexible, loosely connected networks over hierarchical structures, enabling the viral circulation of information and the quick mobilization of people around shared grievances. Participatory propaganda complements this logic by providing the ideological fuel for these aggregated communities.

As individuals coalesce online, they don't just consume content; they actively produce, share, and amplify persuasive messages, reinforcing collective narratives. The ABI movement is a prime example: a viral video served as both a catalyst and a connective tissue, rapidly aggregating individuals and enabling a form of participatory propaganda that influenced national discourse and translated into large-scale physical protests.

The mutually reinforcing relationship between aggregation and participatory propaganda creates a powerful, yet fragile, model for social movements in the digital age. While this combination can lead to rapid mobilization and significant political effects, it often lacks sustainability. The same logic that facilitates swift aggregation can also cause movements to dissolve just as quickly as they form. Without enduring organizational structures or long-term strategies, these movements struggle to maintain momentum beyond short-term peaks. Therefore, while participatory propaganda is highly effective at drawing ordinary users into ideological struggles and activating them in digital spaces, its long-term impact can be ephemeral, highlighting the precarious nature of digitally-driven collective action.

Digital Tactics for Mobilization Within Participatory Propaganda Model

In this section, I discuss six digital tactics the Islamist movement uses to engage online audiences and actively involve them in disseminating persuasive messages, which contributes to successful mobilization. As the table below will demonstrate, the observed mobilization process featured both explicit and implicit modes of participatory propaganda (Repnikova & Fang, 2018), with a predominant use of explicit calls for involvement.

Table 1: Digital Tactics in Participatory Propaganda

Tactic	Key Mechanisms	Examples & Impacts
Provocative Content	Fake news, memes, emotional narratives inciting outrage	Edited video sparked viral outrage; 346K views, 11K shares; catalyzed ABI mobilization
Echo Chambers	Selective information, alternative media, ideological isolation	Exclusive use of Islamic media (e.g., habiebrizieq.com); #BoikotMetroTV or mainstream media boycott; migration to alt-tech platforms like RTimes
Algorithm Manipulation	Use of bots, buzzers, influencers, symbolic hashtags to game platform visibility	Muslim Cyber Army, trending hashtags; human-led manipulation key over automation
Call to Action / Internet Jihad	Narratives invoking Islamic duty to share, like, post; content rewards	Quotes from ulama, Quranic references; youth posting selfies; calls to "use smartphones as weapons"
Crowdsourcing (Beyond Content)	Fundraising, volunteer coordination, logistics via online platforms	MCA recruitment through shahada pledge; Grassroots community logistics; Muslim punk group providing demo support
Act of Disconnection	Unfriending, group exits, polarization in digital and real-world communities	WhatsApp group exits, mosque refusals to accept Ahok supporters' funerals; intensified social and religious polarization

Conclusion

The article has examined the role of digital technologies in the emergence of far-right Islamist protest mobilization in Indonesia, specifically focusing on six digital tactics that contribute to participatory propaganda cascades. Building on previous research, I have suggested that the capabilities of digital technologies play a crucial role in shaping contemporary social movement mobilization. To substantiate this contention, the article moves away from the binary discussions surrounding the rise of far-right Islamist mobilization. Instead, it argues that since the advent of the privately owned social networking sites in Indonesia, the process of Islamist mobilization online has undergone a transformation toward a more participatory form. In contrast to the authoritative, top-down modes of propaganda in traditional Islamist movements, the contemporary mobilization efforts strategically leverage the interactive nature of digital communication, fostering a collaborative effort between far-right activists, political elites, hyper-partisan media, social media influencers, and ordinary citizens.

This abstract was accepted for the conference but has not been presented in person.

References

Jowett, G. S., & O'Donnell, V. (2015). *Propaganda & persuasion* (6th ed.). Sage.

Kydd, A. H. (2021). Decline, radicalization and the attack on the US Capitol. *Violence: An International Journal, 2*(1), 3–23. https://doi.org/10.1177/26330024211010043

Ledford, H. (2024). Deepfakes, trolls and cybertroopers: How social media could sway elections in 2024. *Nature, 626*(7999), 463–464. https://doi.org/10.1038/d41586-024-00314-5

Lewandowsky, S. (2022). Fake news and participatory propaganda. In R. Pohl (Ed.), *Cognitive illusions: Intriguing phenomena in thinking, judgment, and memory* (3rd ed., pp. 324–340). Routledge. https://doi.org/10.4324/9781003154730-23

Rasidi, P. P. (2023). Ludic cybermilitias: Shadow play and computational propaganda in the Indonesian predatory state. *Communication, Culture & Critique, 16*(4), 235–242. https://doi.org/10.1093/ccc/tcad020

Repnikova, M., & Fang, K. (2018). Authoritarian participatory persuasion 2.0: Netizens as thought work collaborators in China. *Journal of Contemporary China, 27*(113), 763–779. https://doi.org/10.1080/10670564.2018.1458063

Starbird, K., DiResta, R., & DeButts, M. (2023). Influence and improvisation: Participatory disinformation during the 2020 US election. *Social Media + Society, 9*(2), 1–17. https://doi.org/10.1177/20563051231179510

Wanless, A., & Berk, M. (2017). Participatory propaganda: The engagement of audiences in the spread of persuasive communications. In D. Herbert & S. Fisher-Høyrem (Eds.), *Social media and social order* (pp. 111–139). De Gruyter. https://doi.org/10.1515/9783110501376-009

Wanless, A., & Berk, M. (2020). The audience is the amplifier: Participatory propaganda. In P. Baines, N. O'Shaughnessy, & N. Snow (Eds.), *The SAGE handbook of propaganda* (pp. 85–104). SAGE. https://doi.org/10.4135/9781526477170

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

AI Narrative Breakdown

A Critical Assessment of Power and Promise of Sustainable AI

Rainer Rehak \\ Weizenbaum Institute \\ Berlin \\ Germany
rainer.rehak@weizenbaum-institut.de

KEYWORDS
artificial intelligence; sustainability; narratives; assessment; power; contestation

Abstract

This article sets off for an exploration of the still evolving discourse surrounding artificial intelligence (AI) in the wake of the release of Large Language Models (LLMs) within the digitalisation and sustainability nexus (Lange & Santarius, 2020). It scrutinizes the pervasive narratives that are shaping the societal engagement with AI (Rehak, 2021), spotlighting key themes such as general purpose, agency and decision-making, autonomy, truthfulness, knowledge processing, prediction, neutrality and objectivity, apolitical optimization and democratization. Those narratives combined fuel the image of AI as "sustainability game-changer" (UN, 2024) forming the basis for fostering democracy and collective well-being. The dualistic portrayal of AI as either a harbinger of societal utopia or dystopia makes it even more difficult to evaluate potentials. This article therefore sets out, to analyse those narratives critically based on insights from critical computer science, critical data and algorithm studies, STS, data protection theory, as well as from the philosophy of mind and semiotics.

To properly analyse those narratives, the article first delves into a historical and technical overview to contextualizes the AI evolution from its inception in 1955 to its current manifestations in both symbolic and subsymbolic systems including the advancement of the AI discourse itself. The article then introduces the notion of "Zeitgeist AI" to critique the imprecise, all-encompassing and misleading use of the term "AI" across various societal sectors including energy, transport and environmental protection. Then, by discussing those narratives with technical expertise and nuance, the article contextualises and challenges often assumed socio-political implications of AI, uncovering in detail and with examples the inherent political, power infused and value-laden decisions within all AI applications. At this point, the reconnection with the sustainability discourse becomes fruitful in understanding the glass ceiling of transformation (Hausknost, 2019) also in regard to AI.

Concluding with a call for a more grounded sustainability engagement with AI, the article carves out acute problems ignored by the narratives discussed, and proposes new socio-technical narratives recognizing AI as a human-directed tool necessarily subject to societal governance and with a need to be integrated into existing democratic structures and sustainability regimes (e.g. Kwet, 2024). The conference presentation was based on two recent publications (Rehak 2024, 2025).

References

Hausknost, D. (2020). The environmental state and the glass ceiling of transformation. *Environmental Politics, 29*(1), 17–37. https://doi.org/10.1080/09644016.2019.1680062

Kwet, M. (2024). Digital degrowth: Technology in the age of survival. Pluto Press.

Lange, S., & Santarius, T. (2020). Smart green world?: Making digitalization work for sustainability. Routledge.

Rehak, R. (2021). The language labyrinth: Constructive critique on the terminology used in the AI discourse. In P. Verdegem (Ed.), *AI for everyone? Critical perspectives* (pp. 87–102). University of Westminster Press. https://doi.org/10.16997/book55.f

Rehak, R. (2024). On the (im)possibility of sustainable artificial intelligence. In T. Züger & H. Asghari (Eds.), *AI systems for the public interest*. *Internet Policy Review, 13*(3). https://doi.org/10.5281/zenodo.14283596

Rehak, R. (2025). AI narrative breakdown: A critical assessment of power and promise. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25)* (pp. 1250–1260). Association for Computing Machinery. https://doi.org/10.1145/3715275.3732083

United Nations Office of the High Commissioner for Human Rights. (2024, June 14). *Artificial intelligence: A game-changer for sustainable development*. https://www.ohchr.org/en/stories/2024/06/artificial-intelligence-game-changer-sustainable-development

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

8 Billion Artists: The Next Generation of Rights in the Next Generation of Generative-AI Creative Works

Zachary Cooper \\ Vrije Universiteit \\ Amsterdam \\ Netherlands
z.b.cooper@vu.nl

KEYWORDS
artificial intelligence; copyright; creativity; intellectual property

Abstract

Generative AI has set copyright frameworks ablaze. Courts worldwide have sought to deny copyright to works developed with generative AI, hoping to enforce a dichotomy of rights in otherwise identical works (GenAI vs non-GenAI) without any reliable means of auditing or enforcement – instead trusting disclosure and non-functional techno-solutions to ascertain when and how AI has been used in a work's production. Yet generative AI is not only transformative in how it allows us to create works. It is changing the forms of the works themselves. Right now, innovators are working tirelessly to ratchet up the level of control and speed in AI generation. This means that we are seemingly on the eve of a new era: one where creative works can be changed in real-time to the whims of the consumer. Just as film cameras and audio recording gear paradigmatically shifted our relationships to media consumption, we are now entering a new phase of creativity. It demands a new phase of regulatory creativity to meet it.

Generative AI has set copyright frameworks ablaze (Burk, 2023; Cooper, 2024; Guadamuz, 2023; Lee, 2024; Lemley, 2024; Hugenholtz and Quintais, 2021). Courts worldwide have sought to deny copyright to works developed with generative AI, hoping to enforce a dichotomy of rights in otherwise identical works (GenAI vs non-GenAI) without any reliable means of auditing or enforcement – instead trusting disclosure and non-functional techno-solutions to ascertain when and how AI has been used in a work's production (Cooper, 2024). Worse, copyright is being separated out granularly within works between those elements which were created using AI and those which weren't, incentivising the development of unreliable surveillant artist-process-tracking technologies. In the mire of this chaos, content platforms are increasingly flooded with generative-AI-assisted works as

algorithms evaluate their popularity to generate and regenerate the best possible works for greater algorithmic visibility.

Critically, generative AI seemingly undermines the conventional economic rationale for copyright – that rational market participants will not invest in the creation of a work if they are not then granted exclusive rights to it – as miniscule investment can produce enormous gluts of content (Hurt and Schuchman, 1966; Ohly and Klippel, 2007; Klass et al., 2021; Drahos and Braithwaite, 2002; Directive 2001/29/EC; U.S. Constitution, Art. I, § 8, cl. 8).

In turn, copyright is at a purposive crossroads – should it seek to stimulate a certain type of higher-investment content, which is not easily generated (say, a huge-cost Hollywood production)? Or does it only need to support the functioning of a market, even if that market is flooded with low-investment content (say, relentless droves of AI slop)? How might it hope to differentiate which should receive rights, and critically – is a rights framework the best means of regulating this?

Yet generative AI is not only transformative in how it allows us to create works. It is changing the forms of the works themselves. Of course, artists can now easily create songs that never end and artworks that continue to expand infinitely, as generative AI shatters former temporal and spatial boundaries. However, these are not the most radical implications of generative AI for creativity. Right now, innovators are working tirelessly to ratchet up the level of control and speed in AI generation.[1] This means that we are seemingly on the eve of a new era: one where creative works can be changed in real-time to the whims of the consumer.

What is the role of copyright in a world where everything can be turned into everything else at the direction of the consumer? Where creative works can be remixed instantaneously into the preferred form of their audiences? Fundamentally, this ability for bi-directional creativity – wherein each artwork becomes its own canvas, each song its own instrument – foundationally reorients our relationship with creativity and consumption, blurring the line between artist and audience. Critically, there is also a bi-directional relationship between copyright itself and the markets it seeks to stimulate. Copyright's role as a "production incentiviser" is especially tenuous as content markets continue to shift towards prosumer platform environments where algorithmic visibility is dependent on huge amounts of content being created and repurposed constantly. These are the exact market dynamics that generative AI allows content creators to meet, who can competitively knock up higher and higher quality videos multiple times a day, complete with bespoke soundtracks and aesthetic differentiation – no longer confined to their present circumstance.

Further, if innovation enables GenAI-enabled *interactive* content to develop as a popular norm for platform prosumption, with content immediately being fed back into the platform to fuel mass viral repurposing, copyright laws predicated on protecting your economic rights through controlling uses of your own work may dwindle in relevance. Rather, overly controlled works may be least valuable for their absence from interactive content platforms actively populated by millions of

1 See, for example, the work of the Dadabots at www.dadabots.com.

prosumers who want to *play* with the content. In turn, copyright's relevancy may (in many instances) be mitigated to a right to "choose" to release your work into the wild and hope to subsidiarily profit.

If the purpose of copyright is inherently challenged then, what role *should* it play in governing emerging digital works, whose forms and functionalities (as well as the markets in which they circulate) radically differ from those that foundational copyright frameworks were developed to regulate? As we enter an age where individuals are able to create infinite streams of interactive fluid media with minimal effort, copyright's incentive structures are distorted, its objectives perverted and (if we are unable to meet this moment) its abuse protected (Cooper, 2024). Just as film cameras and audio recording gear paradigmatically shifted our relationships to media consumption, we are now entering a new phase of creativity. It demands a new phase of regulatory creativity to meet it.

References

Burk, D. L. (2023). Cheap creativity and what it will do. Georgia Law Review, 57.

Cooper, Z. (2024). The AI authorship distraction: Why copyright should not be dichotomised based on generative AI use. Journal of the Copyright Society. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4932612

Cooper, Z., Lehr, W., & Stocker, V. (2024). The new age: Legal & economic challenges to copyright and creative economies in the era of generative AI. The Digital Constitutionalist. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5022340

Dadabots. (n.d.). Dadabots. Retrieved February 14, 2025, from https://www.dadabots.com/

Directive 2001/29/EC. (2001). On the harmonisation of certain aspects of copyright and related rights in the information society. Official Journal of the European Communities, L 167.

Drahos, P., & Braithwaite, J. (2002). Information feudalism: Who owns the knowledge economy. The New Press.

Guadamuz, A. (2023). A scanner darkly: Copyright infringement in artificial intelligence inputs and outputs. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4440733

Hugenholtz, P. B., & Quintais, J. P. (2021). Copyright and artificial creation: Does EU copyright law protect AI-assisted output? IIC—International Review of Intellectual Property and Competition Law, 52, 1193–1213. https://doi.org/10.1007/s40319-021-01117-8

Hurt, R. M., & Schuchman, R. M. (1966). The economic rationale of copyright. The American Economic Review, 56, 425–453.

Klass, N., Stamatoudi, I., & Torremans, P. (2021). Bringing Europe's cultural heritage online: Initiatives and challenges. In I. Stamatoudi & P. Torremans (Eds.), EU copyright law (pp. 959–976). Edward Elgar Publishing.

Lee, E. (2024). The code red for copyright law. Florida Law Review, 76.

Lemley, M. (2024). How generative AI turns copyright law upside down. Science and Technology Law Review, 25.

Ohly, A., & Klipper, D. (2007). Geistiges Eigentum und Gemeinfreiheit (3rd ed., § 11). Mohr Siebeck.

U.S. Constitution.

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

The Critical Participatory Co-Design Method

Exploiting Biases in Generative AI to Grapple with Digital Futures

Regina Sipos \\ Technical University of Munich \\ Munich \\ Germany
regina.sipos@tum.de

KEYWORDS
participation; critical making; generative AI; digital transformation; future scenarios

Abstract

This extended abstract summarizes the Critical Participatory Co-Design (CPC) Method. Responding to the need for broader democratic participation, this method facilitates the participation of communities in envisioning responsible technology design and deployment for large-scale transitions. The CPC method combines Participatory Design's paper prototyping for scenario building and the creation of shared languages, the steps of Critical Making, and exploits biases in generative AI to create reflexive discussions among participants. Examples of how participants interacted with the method are presented to summarize the first steps towards its validation.

Introduction

Science has long been concerned with finding ways to address the societal impact of technologies. Feenberg argued that modern technologies are not neutral, and their negative effects are not mere unintended side effects. He advocated for broader societal participation to understand technology's democratic potential (Feenberg 1991; 1999; 2005). As humanity continues to drive large-scale digital transformations, it will remain important to critically examine and reflect on the promises, imaginaries, and biases that such endeavors reproduce and amplify.

This study focuses on the Critical Participatory Co-Design (CPC) method (Sipos, 2025). The method was developed to support reflexive participation utilizing speculative methods. It allows broader

audiences to engage with "postdigital futures," i.e., what might lie beyond the hype and promises of digital solutions (Macgilchrist et al. 2024).

The quick development and general accessibility of generative artificial intelligence (genAI) create a window of opportunity to utilize the technology in novel ways. Learning from Participatory Design methods, shared languages must be created between diverse participants to generate visions of more democratic technological futures (Bødker et al. 2000). Recent approaches in Transition Design focus on mapping systems with and showing stakeholders common starting points for interventions (Irwin 2018), supporting them to focus on the changeable. Furthermore, with complexity, quick fixes are not sufficient (Barendregt et al. 2024). Complex issues, also known as wicked problems (Rittel and Webber 1973), require long-term involvement, but such commitment excludes many members of society. The question of who gets to participate (Elovaara et al. 2006; Macgilchrist et al. 2024) and how equitable participation can be enabled is to be critically examined if we want visions of possible futures to include broader societal considerations.

The CPC Method in Brief

Before this background, the CPC method addresses the following three questions: how to create shared languages between diverse participants; how to support critical thinking and reflection about deeply political sociotechnical questions in an accessible manner and through these, the imagination of alternative futures; and how to achieve this in a limited amount of time, allowing for more democratic participation.

One example of a major and upcoming arena of digital transformation affecting many is the transition of smart cities to metacities (Bibri et al. 2022). It promises that all infrastructure related to a city – education, bureaucratic processes, retail, and social life – will be fully virtual. Yet, our struggle with technology-driven, sterile, and unappealing visions of the smart city was already questioned a decade ago (Greenfield 2013), lacking a responsible examination of privacy, inclusion, or access. According to the current sociotechnical imaginary, the shortcomings of smart cities can be redeemed with new transitions in a people-centered way (ITU 2023).

Thus, a workshop organized for urban designers interested in this topic provided an adequate backdrop to test the CPC method. Before the workshop, expert validation of the method was conducted (n=6), and during the workshop with urban designers (n=15), empirical research was conducted to collect qualitative and comparative quantitative data from pre- and post-workshop surveys.

Briefly summarized for this extended abstract, the CPC method combines three methods. First, paper prototyping, a tool used in Participatory Design to create shared languages (Simonsen and Robertson 2013). Second, critical making, an interactive method developed to unpack wicked problems (Ratto and Hockema 2009). Third, human-AI collaboration, image-to-image genAI is used for the exploitation of biases inherent in AI datasets, as visual starting points for reflexive discussions. This

step is facilitated with an open-source device built for the workshop, called Street2Paradise_[1] that is capable of taking photos of the scenarios, a simple user interface for prompts, and a screen to show the AI's interpretation.

The workshop follows the constructivist steps of critical making (ibid.): research, making, and reflection. The making step includes analog mock-ups of scenarios as agreed upon by the participants, depicting what various events and spaces might look like in the metaverse. Then, a speculative genAI visualization is created with a low-cost open-source machine to support the critical-speculative exploration of digital futures. The practical part is supplemented with reflexive questions regarding stereotypes reproduced by the genAI, inclusion, and exclusion in future visions, and an actionable sharing session on how the participants might use the CPC method.

Scenarios and Biases Exploited

The participants decided to explore Lively Neighborhoods, Churches as Community Spaces, Higher Education, and Spaces of Celebration in the Metacity as scenarios (see Figure 1).

Figure 1: The 4 paper mock-up scenarios and their AI interpretations

Lively Neighborhoods

The first group asked themselves what a lively and inclusive neighborhood might look like in a Metacity in the year 2100. If all our activities take place in the virtual space, including education, bureaucratic processes, retail, and social life, and we might not need any physical infrastructure anymore outside of our homes, can we still create lively and inclusive neighborhoods to meet?

In terms of the biases exploited for reflection, the first group highlighted that when they wanted to create the feeling of a community by adding the terms love and empathy to their prompts – and the genAI responded with feminine-looking figures. As a group member said, it seemed as if "these skills were uniquely reserved for women".

1 Documentation can be found under https://github.com/vektorious/street2paradise.

Churches as Community Spaces

The second group built on the assumption that in the Metacity, people will still want to meet in person to exercise their religion. Inspired by a current project of one of the participants, the group decided to envision churches in the future as places of community, gathering, and exchange.

Reflecting on their collaboration with the AI, second group highlighted two major points of frustration: in their interpretation, the church was repeatedly depicted as a "historically anchored" building, and they had difficulty moving away from the values seemingly embedded in the dataset; yet, when they added the term activism, they did not see any interaction or transaction between the activists sitting at the table in the forefront and the church representatives standing in the back. The group wished for less normativity and more creativity detached from what, in their eyes, were historical-traditional visions for the church's role.

Higher Education

The third group had two participants working with the educational sector, one of them working on inclusive, multicultural spaces. Thus, they asked themselves, what would higher education look like in a Metacity?

The third group evaluated their outcomes as very cliché-like, resembling futuristic artworks from the 1980s, with white, Saturn-like planets flying in the sky. Then, the term poverty was added to consider what education in a low-income setting might look like in a Metacity. This resulted in an image that, according to their evaluation, reproduced a major stereotype: a person either in a blue headscarf or a hoodie, indicating that the biases we currently hold in the West about such attire will continue to be present in the future.

Spaces of Celebration

The fourth group wondered what a social event might look like in the Metacity. After an initial discussion on how family celebrations were restricted during the COVID-19 pandemic, they settled on children's birthdays to better understand how a birthday could be celebrated in the metaverse if another lockdown would happen in the future.

The fourth group criticized their genAI companion for not being very creative with the future, not thinking beyond concrete in the cityscape, and creating too clean visuals; for example, no graffiti was to be seen in the pictures. The outcomes were evaluated as abstract rather than concrete, idyllic and harmonic, resembling "crazy dreams" rather than reality or future visions.

Preliminary Validation

Preliminary data shows the validity of the workshop in the three considerations outlined above. The participants positively evaluated both critical making and paper prototyping to create shared languages between them. The participants appreciated critical making as an approach to "not being always so cerebral, yet receiving stimulus for reflection", supporting critical thinking and reflection in an accessible way. Finally, the statistical analysis of the pre- and post-workshop surveys, using a paired t-test showed a significant increase in how often participants plan to use critical thinking in their work after the workshop (p-value =0.026).

Most importantly, participants declared themselves ready to use the method as a tool in their own participatory praxis: to generate ideas with the public, to engage youth in urban design workshops and school classes as a critical educational tool supporting critical discussions among students, and to teach pupils media literacy skills. In summary, they evaluated the method as applicable for participatory engagement with broader audiences.

Conclusion

To conclude, this bricolage of methods could be used in any setting where large-scale digital transformation is planned, and decision-makers wish to handle this responsibility in a democratic and anticipatory manner. Visions of possible futures contributing to large-scale digital transformations need to include broader societal considerations enabled by validated methods supporting equitable participation. CPC is thus a method to reopen Feenberg's problem space of broader democratic participation in digital transitions, contributing to discussions on how to empower people in online spaces and better represent citizens' perspectives.

Acknowledgement

I would like to wholeheartedly thank Dr. Alexander Kutschera for the fantastic collaboration over the last years: for developing the Street2Paradise_ device and for co-moderating the first workshop.

References

Barendregt, L., Bendor, R., & Van Eekelen, B. F. (2024). Public participation in futuring: A systematic literature review. *Futures, 158*, 103346. https://doi.org/10.1016/j.futures.2024.103346

Bibri, S. E., Allam, Z., & Krogstie, J. (2022). The metaverse as a virtual form of data-driven smart urbanism: Platformization and its underlying processes, institutional dimensions, and disruptive impacts. *Computational Urban Science, 2*(1), 24. https://doi.org/10.1007/s43762-022-00051-0

Bødker, S., Ehn, P., Sjögren, D., & Sundblad, Y. (2000). Co-operative design — Perspectives on 20 years with 'the Scandinavian IT design model'.

Elovaara, P., Igira, F. T., & Mörtberg, C. (2006). Whose participation? Whose knowledge? Exploring PD in Tanzania-Zanzibar and Sweden. In *Proceedings of the Ninth Conference on Participatory Design: Expanding Boundaries in Design* (Vol. 1, p. 105). ACM Press. https://doi.org/10.1145/1147261.1147277

Feenberg, A. (1991). *Critical theory of technology*. Oxford University Press.

Feenberg, A. (1999). *Questioning technology*. Routledge.

Feenberg, A. (2005). Critical theory of technology: An overview. *Tailoring Biotechnologies, 1*(1), 47–64.

Greenfield, A. (2013). *Against the smart city*. Do Projects.

Irwin, T. (2018). The emerging transition design approach. In C. Storni, K. Leahy, M. McMahon, P. Lloyd, & E. Bohemia (Eds.), *Design as a catalyst for change – DRS International Conference 2018, 25–28 June, Limerick, Ireland*. https://doi.org/10.21606/drs.2018.210

International Telecommunication Union. (2023). *ITU webinars: Digital transformation. Episode 35. How to successfully develop people-centred citiverse?* https://www.itu.int/cities/standards4dt/ep35/

Macgilchrist, F., Allert, H., Cerratto Pargman, T., & Jarke, J. (2024). Designing postdigital futures: Which designs? Whose futures? *Postdigital Science and Education, 6*(1), 13–24. https://doi.org/10.1007/s42438-022-00389-y

Ratto, M., & Hockema, S. (2009). FLWR PWR – Tending the walled garden. In A. Dekker & A. Wolfsberger (Eds.), *Walled Garden* (pp. 51–62). Virtueel Platform.

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences, 4*(2), 155–169.

Simonsen, J., & Robertson, T. (Eds.). (2013). *Routledge international handbook of participatory design*. Routledge.

Sipos, R. (2025). Using generative AI to grapple with the not-yet-known: A case study on the critical participatory co-design method. In *Proceedings of the 12th International Conference on Communities & Technologies (C&T 2025), July 20–23, 2025, Siegen, Germany* (13 pages). ACM. https://doi.org/10.1145/3742800.3742844

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Promoting Computational Empowerment for Youth

Designing and Auditing Generative AI Filters in Social Media

Yasmin B. Kafai \\ University of Pennsylvania \\ Philadelphia, PA \\ USA
kafai@upenn.edu

Luis Morales-Navarro \\ University of Pennsylvania \\ Philadelphia, PA \\ USA

Danaé Metaxa \\ University of Pennsylvania \\ Philadelphia, PA \\ USA
metaxa@upenn.edu

KEYWORDS
algorithm auditing; computational empowerment; social media; generative AI

Abstract

The increasing presence of AI and machine learning in young people's lives underscores the need for computational empowerment in education. This paper explores algorithm auditing as a method to engage youth in critically analyzing AI systems. Through participatory workshops, high school youth audited TikTok's AI-powered filters, identifying biases in racial and gender representation. Using a structured five-step model, youth systematically tested AI outputs, linking biases to design choices. Findings demonstrate that youth can effectively conduct algorithm audits, fostering critical awareness of AI systems. We discuss implications for integrating auditing into classrooms to enhance computational literacy and ethical AI engagement.

Introduction

The increasing presence of AI and machine learning (AIML) technologies in young people's lives — social media filters, recommendation systems, voice assistants, and more — has led to calls for enhancing youths' understandings of these applications (e.g., Long & Magerko, 2020; Touretzky et al., 2019). Several frameworks, in particular computational empowerment, encourage learners to

"engage critically and curiously with the construction and deconstruction of technology" (Dindler et al., 2022, p. 121), addressing both functional and ethical issues (DiPaola et al., 2020). However, integrating this approach into K–12 education poses challenges, including the need for participatory practices, educational principles, and scaffolding tools.

In this paper we propose the use of algorithm auditing as a method to foster computational empowerment in youth. Algorithm auditing (also called AI auditing) involves "repeatedly querying an algorithm and observing its output to draw conclusions about the algorithm's opaque inner workings and potential external impacts" (Metaxa et al., 2021b, p. 272). Auditing is traditionally a practice conducted entirely by experts (Metaxa et al., 2021a; Sandvig et al., 2014; Sweeney, 2013), though more recent work has begun including everyday users as participants in audits (DeVos et al., 2022; Lam et al, 2020; Shen et al., 2021). It involves systematically testing AI systems to reveal biases and social impacts. Prior research has involved youth in software testing (Druin, 2002), but the potential for youths' involvement in systematic AI auditing remains underexplored (e.g., Zimmerman-Niefeld et al., 2020). Our research focused on engaging high school youth in participatory workshops designing (constructing) and auditing (deconstructing) AI-powered image filters used by the video-based social media platform TikTok. By piloting instructional scaffolds for auditing, we helped youth systematically analyze AI's impact and biases (Coenraad, 2022; Salac et al., 2023; Solyst et al., 2023), and moreover to identify specific issues of interest to the youths themselves. Algorithm auditing expands existing AI education efforts by positioning youth as external evaluators.

Context

We conducted workshops at a local science center with groups of high school youth (ages 14-16 years) to teach them about algorithm auditing and understand how they engaged in the process. Adapting expert auditing methods (Morales-Navarro et al., 2025) into a five-step model — hypothesis development, input generation, testing, analysis, and reporting — created a structured, accessible approach for youth (see Table 1). The specific AI model being audited was the text-to-image AI model available through Effect House, TikTok's AI filter development environment. This model applies the user's text prompt (e.g., "anime style") to any input image, also provided by a user (e.g., a photo of a celebrity), outputting a stylized version of the original (e.g., an image that looks similar to the input celebrity photo, but stylized in the anime illustration style).

Table 1: Algorithm Auditing Approach: Five Steps and Scaffolds

Activity/Step	Scaffolds
Step 1: Developing a Hypothesis	Provided examples of hypotheses, modeled how to generate a hypothesis together, time for open-ended exploration or play.
Step 2: Generating a set of systematic, thorough, and thoughtful inputs	Designed a two-axis input organizer and asked the youth to fill it out with at least 30 images
Step 3 – Running tests	Designed a table for students to record input and output pairs and take notes on specific things in their outputs
Step 4 – Analyzing the data	Suggested ideas for the kind of analysis the students could do (calculating percentages/describing in detail the outputs)
Step 5 – Reporting results	Recommended potential audiences and gave them examples of audit reports done in previous workshops

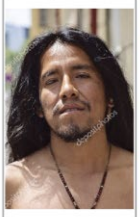
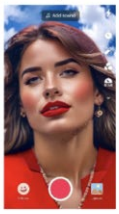
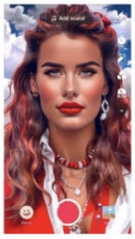
Findings

We present a case study following two youths in our workshop, Ishmael (14 years) and Ziyi (15 years), as they designed and audited generative AI filters using TikTok's Effect House. In terms of designing AI filters, Ishmael and Ziyi experimented with prompts to create AI-generated effects starting with broad descriptions (e.g., "elephant, trunk, grey"), but found the changes to their input images were minimal. Adjusting the prompt strength (a numerical parameter determining the degree to which the filter is allowed to change the input image) improved their transformations. They then designed a "Disney Princess Aurora" filter that they hoped would make their outputs look like a Disney princess version of the input images, but were disappointed to find that the filter consistently lightened input images' skin tones. This led them to create a "Princess Tiana" filter with attributes specific to that African American character. Their testing revealed that it darkened all skin tones. The combination of their testing of these two filters prompted discussions about racial bias in AI-generated imagery.

Next, our activity had them change perspectives, from designers of AI filters and testers of their own design, to external auditors of their peers' design. Without access to the input prompt used by their peers, youth audited a peer-designed filter that gave users red hair, red clothing, and a cloudy sky background. They hypothesized that it exaggerated feminine traits and tested this hypothesis on a diverse set of images. Despite their test dataset including photos of men, women, and non-binary individuals of various racial backgrounds, they observed that the filter's outputs all appeared

female, with many displaying Eurocentric features like blush and tanned skin (see Figure 1). They systematically analyzed the results, documenting and counting hairstyle, clothing, and gender presentation changes. Finally, they shared their findings in a humorous TikTok video, energetically and humorously explaining their audit protocol and its conclusions. Through this process, Ishmael and Ziyi critically explored AI biases, particularly those pertaining to racial and gender representation, while learning about prompt design and systematic auditing.

Figure 1: Auditing the filter: An excerpt of users' table for recording audit results containing input and output pairs along with their notes.

Input	Output	Hair	Gender	Skin
		Hair stayed wavey. Hair color became brown with red highlights.	M to F	Light tan skin turned to orange tan skin
		Hair stayed curly. Hair become red with blone highlight Highlight on hair stayed	M to F	Skin tone stayed the same Makeup was added.
		The hair stayed curly The color looks very similar but became brighter. Turnmed blonde highlight into a red highlight	M to F	Skin tone become more tanned.

Discussion

This paper investigates algorithm auditing as a method to empower youth in critically analyzing and evaluating machine learning (ML) applications. Youth followed a five-step model guiding their algorithm audits of a TikTok filter. The model was based on literature on expert- and end user-driven algorithm audits, helping them make inferences about AI models' data and design. Youth identified

biases and linked them to specific design choices, demystifying ML systems. Related research demonstrated that algorithm auditing of youth was as effective in identifying algorithmic bias in social media filters as that of experts, even though they used smaller data sets for their audits and analysis.

To enhance auditing activities for youth, we suggest improving support for data collection, analysis, and reporting. Generating datasets collaboratively and incorporating statistical tools for analysis could make the process more effective. Future implementations might include a sixth step focusing on their reflections. Additionally, integrating auditing into classrooms requires selecting AI/ML systems suitable for the classroom, establishing collaborative structures, and training teachers. Limited access to TikTok in many schools is a challenge that highlights the need for alternative applications and settings. Ultimately, we position algorithm auditing as a powerful tool for computational empowerment, helping youth critically engage with AI/ML technologies while fostering deeper understanding and awareness of biases and system design choices.

Funding Information and Acknowledgements

The work reported in this paper was supported by National Science Foundation grants #2333469 and #2342438 to the first and last authors. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation or the University of Pennsylvania.

References

Coenraad, M. (2022). "That's what techquity is": Youth perceptions of technological and algorithmic bias. *Information and Learning Sciences, 123*(7/8), 500–525. https://doi.org/10.1108/ILS-02-2022-0023

DeVos, A., Dhabalia, A., Shen, H., Holstein, K., & Eslami, M. (2022, April). Toward user-driven algorithm auditing: Investigating users' strategies for uncovering harmful algorithmic behavior. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1–19). ACM. https://doi.org/10.1145/3491102.3501821

Dindler, C., Smith, R., & Iversen, O. S. (2020). Computational empowerment: Participatory design in education. *CoDesign, 16*(1), 66–80. https://doi.org/10.1080/15710882.2019.1654526

DiPaola, D., Payne, B. H., & Breazeal, C. (2020, June). Decoding design agendas: An ethical design activity for middle school students. In *Proceedings of the Interaction Design and Children Conference* (pp. 1–10). ACM. https://doi.org/10.1145/3392063.3394433

Druin, A. (2002). The role of children in the design of new technology. *Behaviour & Information Technology, 21*(1), 1–25. https://doi.org/10.1080/01449290110108659

Lam, M. S., Gordon, M. L., Metaxa, D., Hancock, J. T., Landay, J. A., & Bernstein, M. S. (2022). End-user audits: A system empowering communities to lead large-scale investigations of harmful algorithmic behavior. *Proceedings of the ACM on Human-Computer Interaction, 6*(CSCW2), 1–34. https://doi.org/10.1145/3555101

Long, D., & Magerko, B. (2020, April). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). ACM. https://doi.org/10.1145/3313831.3376727

Metaxa, D., Gan, M. A., Goh, S., Hancock, J., & Landay, J. A. (2021). An image of society: Gender and racial representation and impact in image search results for occupations. *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW1), 1–23. https://doi.org/10.1145/3449102

Metaxa, D., Park, J. S., Robertson, R. E., Karahalios, K., Wilson, C., Hancock, J., & Sandvig, C. (2021). Auditing algorithms: Understanding algorithmic systems from the outside in. *Foundations and Trends® in Human–Computer Interaction, 14*(4), 272–344. https://doi.org/10.1561/1100000071

Morales-Navarro, L., Shah, M., & Kafai, Y. B. (2024, March). Not just training, also testing: High school youths' perspective-taking through peer testing machine learning-powered applications. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education, Vol. 1* (pp. 881–887). ACM. https://doi.org/10.1145/3545945.3569821

Morales-Navarro, L., Kafai, Y. B., Vogelstein, L., Yu, E., & Metaxa, D. (2025). Learning about algorithm auditing in five steps: Scaffolding how high school youth can systematically and critically evaluate machine learning applications. In *Proceedings of the AAAI Conference on Artificial Intelligence, 39*.

Salac, J., Landesman, R., Druga, S., & Ko, A. J. (2023, June). Scaffolding children's sensemaking around algorithmic fairness. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference* (pp. 137–149). ACM. https://doi.org/10.1145/3585088.3589370

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). Auditing algorithms: Research methods for detecting discrimination on internet platforms. In *Data and discrimination: Converting critical concerns into productive inquiry* (pp. 43–49). https://doi.org/10.2139/ssrn.2477598

Shen, H., DeVos, A., Eslami, M., & Holstein, K. (2021). Everyday algorithm auditing: Understanding the power of everyday users in surfacing harmful algorithmic behaviors. *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW2), 1–29. https://doi.org/10.1145/3479618

Solyst, J., Yang, E., Xie, S., Ogan, A., Hammer, J., & Eslami, M. (2023). The potential of diverse youth as stakeholders in identifying and mitigating algorithmic bias for a future of fairer AI. *Proceedings of the ACM on Human-Computer Interaction, 7*(CSCW2), 1–27. https://doi.org/10.1145/3610184

Sweeney, L. (2013). Discrimination in online ad delivery: Google ads, black names and white names, racial discrimination, and click advertising. *Queue, 11*(3), 10–29. https://doi.org/10.1145/2460276.2460278

Touretzky, D., Gardner-McCune, C., Martin, F., & Seehorn, D. (2019, July). Envisioning AI for K-12: What should every child know about AI? In *Proceedings of the AAAI Conference on Artificial Intelligence, 33*(01), 9795–9799. https://doi.org/10.1609/aaai.v33i01.33019795

Zimmermann-Niefield, A., Polson, S., Moreno, C., & Shapiro, R. B. (2020, June). Youth making machine learning models for gesture-controlled interactive media. In *Proceedings of the 19th Annual ACM Interaction Design and Children Conference* (pp. 63–74). ACM. https://doi.org/10.1145/3392063.3394406

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Measuring Generative AI Knowledge

A Comparative Study of Self-Assessment, Conceptual Understanding, and Factual Knowledge

Kiran Kappeler \\ University of Copenhagen \\ Copenhagen \\ Denmark
kkap@hum.ku.dk

Michael V. Reiss \\ Leibniz Institute for Media Research | Hans-Bredow-Institut \\ Hamburg \\ Germany
m.reiss@leibniz-hbi.de

Judith Möller \\ Leibniz Institute for Media Research | Hans-Bredow-Institut \\ Hamburg \\ Germany
j.moeller@leibniz-hbi.de

KEYWORDS
generative AI, generative AI knowledge, digital skills, online survey, measurement, digital inequality

Introduction

Since late 2022, generative AI (GenAI) tools like ChatGPT have rapidly gained popularity, with 44% of people in Germany using them (Reiss et al. 2025). Digital literacy, often seen as essential for the successful adoption of digital technologies (Hargittai and Micheli 2019), has been linked to various outcomes, such as effective use of voice assistants (Gruber et al. 2021), privacy protection (Büchi, Just, and Latzer 2017), and managing algorithmic risks (Kappeler et al. 2023). While much research has focused on digital skills (see e.g., Allmann and Blank 2021; Dogruel et al. 2021), the emerging field of GenAI knowledge is still in early stages (see e.g., Mansoor et al. 2024; Pinski and Benlian 2023). This study is grounded in theories of digital divides and inequalities (van Dijk 2020), which suggest that differences in digital literacy can reinforce social ones (Blank and Lutz 2018). We argue that GenAI knowledge, like AI literacy (Wang, Rau, and Yuan 2023), influences the use of GenAI tools, with potential consequences for digital inequalities. Thus, examining GenAI knowledge is crucial as these tools become more widespread in daily life. In this study, we aim at exploring efficient ways of measuring GenAI knowledge, comparing self-assessments, conceptual understanding, and factual

knowledge. Hence, this study offers a reliable scale for measuring GenAI knowledge in future surveys and contributes to the understanding of GenAI knowledge and digital inequalities in the context of human and artificial intelligence.

Method

Data collection for this study was based on an online survey and included a pretest (n = 24) as well as the main data collection, which was conducted in two phases in spring 2025 (phase 1: n = 290; phase 2: n = 787). The sample is representative of internet users in Germany aged 16 years and above regarding age, gender, education and state of residence.

In our analysis, we compared three measures of GenAI knowledge: self-reported GenAI knowledge ("To what degree do you know about AI and chat-based AI applications?" with answer options ranging from 1 = "very bad" to 5 = "very well"), understanding of GenAI-related terms, i.e., GenAI skills (e.g., "prompt", "ChatGPT", with answers ranging from 1 = "I do not understand what is meant by this at all" to 5 = "I fully understand what is meant by this"), and factual GenAI knowledge ("What do you think about the following statements. Are these correct or false?" e.g., "The development of chat-based AI tools requires big quantities of data").

For item development, we followed a multi-step procedure (Wang, Rau, and Yuan 2023; Hinkin 1998). First, we generated the items and tested the general applicability and fit through a pretest with a diverse sample. For the GenAI-related terms, we applied exploratory factor analysis (EFA) based on phase 1 of the data collection and, subsequently, a confirmatory factor analysis (CFA) based on phase 2 of the data collection. The GenAI knowledge scale was developed and validated using the full main sample, based on principles of Item Response Theory—specifically, the Rasch model.

We developed our items based on existing literature on digital and AI literacy skills (see e.g., Dogruel, Masur, and Joeckel 2022; Hargittai et al. 2020; Kappeler 2024; Wang, Rau, and Yuan 2023; Yuan, Tsai, and Chen 2024) and iterative brainstorming within the research team. This process resulted in 14 GenAI-related terms and 23 GenAI knowledge questions.

Preliminary Findings

The EFA in phase 1 led us to nine items loading onto one factor, with factor loadings of at least .56 each. Subsequently, in phase 2, after stepwise exclusion of four misfitting items, model fit was good ($\chi^2(5) = 20.528$, $p < .001$; CFI = .985; RMSEA = .063; SRMR = .025; Cronbach's $\alpha = .79$), meeting Hu and Bentler's (1999) criteria and exceeding the strictest dynamic fit cutoffs per McNeish and Wolf (2023). The five remaining GenAI-related terms are "Generative AI", "OpenAI", "Natural Language Processing", "Dall-E", and "Machine Learning", with loadings of .65 on average and at least .51. Participants' average self-reported understanding of the GenAI-terms was 2.61 (SD = 0.97), while the single-item self-assessment averaged 3.08 (SD = 1.27). Convergent validity was supported by a

significant positive correlation between the two measures ($r = .66$, $p < .001$), indicating they largely assess the same construct. For practical reasons, future researchers might consider using the single-item self-report as an efficient alternative to the 5-item scale on the understanding of GenAI terms. Results for the factual GenAI knowledge items and the Rasch model are currently still outstanding.

Conclusion

This article set out to find out how GenAI knowledge can be measured and how three measurements of GenAI knowledge relate to one another. The comparison revealed a strong correspondence between self-reported GenAI knowledge and participants' self-assessed understanding of GenAI-related terms. Results for the factual GenAI knowledge measure are still pending. Our findings offer valuable insights and guidance on different ways to measure GenAI knowledge, which can be used for future research. Based on the results, using the single-item self-reported measure is often an efficient and appropriate method for broader studies investigating the relationship between GenAI knowledge and other concepts.

References

Allmann, K., and Blank, G. (2021). "Rethinking Digital Skills in the Era of Compulsory Computing: Methods, Measurement, Policy and Theory." *Information, Communication & Society* 24 (5): 633–48. https://doi.org/10.1080/1369118X.2021.1874475

Blank, G., and Lutz, C. (2018). "Benefits and Harms from Internet Use: A Differentiated Analysis of Great Britain." *New Media & Society* 20 (2): 618–40. https://doi.org/10.1177/1461444816667135

Büchi, M., N. Just, and Latzer, M. (2017). "Caring Is Not Enough: The Importance of Internet Skills for Online Privacy Protection." *Information, Communication & Society* 20 (8): 1261–78. https://doi.org/10.1080/1369118X.2016.1229001

Dijk, J. van (2020). *The Digital Divide*. Cambridge, UK; Medford, MA: Polity

Dogruel, L., P. Masur, and Joeckel, S. (2022). "Development and Validation of an AlgorithmLiteracy Scale for Internet Users." *Communication Methods and Measures* 16 (2): 115-133. https://doi.org/10.1080/19312458.2021.1968361

Gruber, J., E. Hargittai, G. Karaoglu, and Brombach, L. (2021). "Algorithm Awareness as an Important Internet Skill: The Case of Voice Assistants." *International Journal of Communication* 15 (0): 19

Hargittai, E., J. Gruber, T. Djukaric, J. Fuchs, and Brombach, L. (2020). "Black Box Measures? How to Study People's Algorithm Skills." *Information, Communication & Society* 23 (5): 764–75. https://doi.org/10.1080/1369118X.2020.1713846

Hargittai, E., and Micheli, M. (2019). "Internet Skills and Why They Matter." In *Society and the Internet*, by E. Hargittai and M. Micheli, 109–24. Oxford University Press. https://doi.org/10.1093/oso/9780198843498.003.0007

Hinkin, T. R. (1998). "A Brief Tutorial on the Development of Measures for Use in Survey Questionnaires." *Organizational Research Methods* 1 (1): 104–21. https://doi.org/10.1177/109442819800100106

Hu, L., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6* (1), 1–55. https://doi.org/10.1080/10705519909540118

Kappeler, K. (2024). "A Longitudinal Perspective on Digital Skills for Everyday Life: Measurement and Empirical Evidence." *Media and Communication* 12 (0). https://doi.org/10.17645/mac.8159

Kappeler, K., N. Festic, M. Latzer, and Rüedy, T. (2023). "Coping with Algorithmic Risks: How Internet Users Implement Self-Help Strategies to Reduce Risks Related to Algorithmic Selection." *Journal of Digital Social Research* 5 (1): 23–47. https://doi.org/10.33621/jdsr.v5i1.130

Mansoor, H. M. H., A. Bawazir, M. A. Alsabri, A. Alharbi, and Okela, Abdelmohsen H. (2024). "Artificial Intelligence Literacy among University Students—a Comparative Transnational Survey." *Frontiers in Communication* 9 (October):1478476. https://doi.org/10.3389/fcomm.2024.1478476

McNeish, D., and Wolf, M. G. (2023). Dynamic fit index cutoffs for confirmatory factor analysis models. *Psychological Methods*, *28* (1), 61–88. https://doi.org/10.1037/met0000425

Pinski, M., and Benlian, A. (2023). "AI Literacy – Towards Measuring Human Competency in Artificial Intelligence." In *56th Hawaii International Conference on System Sciences*. Hawaii. https://hdl.handle.net/10125/102649

Reiss, M. V., Knor, E. L., Stöwing, E., Merten, L., and Möller, J. (2025). Zwischen Neugier und Skepsis: Nutzung und Wahrnehmung generativer KI zur Informationssuche in Deutschland. (Arbeitspapiere des Hans-Bredow-Instituts, 76). Hamburg: Verlag Hans-Bredow-Institut. https://doi.org/10.21241/ssoar.100907

Wang, B., P.-L. P. Rau, and Yuan, T. (2023). "Measuring User Competence in Using Artificial Intelligence: Validity and Reliability of Artificial Intelligence Literacy Scale." *Behaviour & Information Technology* 42 (9): 1324–37. https://doi.org/10.1080/0144929X.2022.2072768

Yuan, C. W. (T.), H.-Y. S. Tsai, and Chen, Y-T. (2024). "Charting Competence: A Holistic Scale for Measuring Proficiency in Artificial Intelligence Literacy." *Journal of Educational Computing Research* 62 (7): 1675–1704. https://doi.org/10.1177/07356331241261206

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Panel 2

Session 4: Power Imbalances and Contestation

Session 5: Political Communication and Social Media

Session 6: Protecting and Building Identities

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Three Contextual Problem-Solving Strategies for Digital Sovereignty as an Ill-Structured Problem

Christian Herzog \\ University of Lübeck \\ Lübeck \\ Germany
christian.herzog@uni-luebeck.de

Robin Preiß \\ University of Lübeck \\ Lübeck \\ Germany
robin.preiss@uni-luebeck.de

Renée Ridgway \\ Aarhus University \\ Aarhus \\ Denmark
rridgway@cc.au.dk

Daniela Zetti \\ Archives of Contemporary History \\ Zurich \\ Switzerland
zettid@ethz.ch

KEYWORDS
digital sovereignty; ill-structured problems; contextual problem-solving; digital transformation; case studies

Abstract

This article addresses ill-structured problems regarding the discourses encircling the term "digital sovereignty," introducing the specific actors (or stakeholders) involved in three European case studies of digitalization, datafication, and internet infrastructures. We employ a case study method of analysis to interweave published talks, a group discussion, and an ethnographical participant observation to reflect on the question of digital sovereignty as an ill-structured problem. By identifying and explaining digital sovereignty through contextual lenses (national, state, individual, federated), it opens up the potential for putting forth well-structured problems. Well-structured solutions for digital sovereignty take form only after various iterations, knowledge exchanges, and practices are clarified, which includes the views of an "emancipated citizenry." These can appear through the design of open source and federated alternatives, as well as through the inclusion of those formally marginalized by society. By demonstrating how, during the past decade and a half, the terminologies surrounding digital sovereignty have been transformed from traditional nation-state

policy-making and individual concerns, it puts forth the advancement of federated infrastructures. The contribution combines perspectives from policy and philosophy of technology by analyzing ill-structuredness in three cases and suggests contextual problem-solving strategies to tackle and further develop digital sovereignty.

Digital Sovereignty as an Ill-Structured Problem?

Digital sovereignty has risen to a key concept indicating both the need for national regulation of digital technologies and the hope to enhance citizens' individual and local sovereignty towards data regimes and global digital infrastructures (Calderaro & Blumfelde, 2022; Couture & Toupin, 2019; Floridi, 2020; Hutchinson et al., 2024; see, e.g., Pohle & Thiel, 2021). Even though digital sovereignty envelops a rhetoric of potential benefits, its implementation also instantiates manifold challenges and complexities that have to be considered, for example, how to balance the legitimate interests of the state's and individuals' data flows. However, policymakers and observers have repeatedly pointed out that the notion of digital sovereignty lacks a clear-cut definition (Pohle & Thiel, 2020). The discursive landscape can be characterized as driven by competing agendas to enhance either individual or territorial sovereignty, attempting to foster participation in complex socio-technical settings.

Thus, Herzog, Zetti, and Preiß (2024) have called for acknowledging that digital sovereignty could be understood as an ill-structured problem (ISP) (cf. Simon, 1973). ISPs require epistemic iterations to promote even an understanding of the issues to solve––in an attempt to ultimately arrive at a clearer, potentially even formalized (well-structured) description. However, we emphasize a need to also take critical note of the power dynamics at play and, therefore, depart from an idealized conception of attempts to tackle ISPs by well-ordered discourse. For instance, narratives brought forward by state actors within public discourse may misappropriate tensions between the individual and territorial conceptions of digital sovereignty because of their one-sided focus, thereby tending to privilege mechanisms that foster territorial sovereignty.

Viewing Digital Sovereignty Through Three Contextual Lenses

We investigate dynamics and tensions such as these by means of a case study method of analysis (cf. Paris et al., 2024) that interweaves published talks, a group discussion, and an ethnographical participant observation to reflect on the question of digital sovereignty as an ill-structured problem, in regard to datafication and infrastructures. Involving scholars, policy makers, public service IT managers, social scientists, and the public (users of platforms), three case studies each identify digital sovereignty as an 'ill-structured' problem, which specifically displays tensions between national, individual, and federated sovereignty. The first case study critically analyzes public

discourses framing digital sovereignty as a (supra-)national matter through the German president's initiative on digital transformation ("Forum Bellevue on the Transformation of Society"). The second case study focuses on a German administrative online platform (ZuFiSH) that might foster citizens' digital sovereignty vis-à-vis administrative technology, analyzing empirical material stemming from a group discussion in 2024. The third considers digital sovereignty as a core value for a forthcoming EU-funded open web index (OWI)[1] through five questions, drawing on answers and reflections from the Working Group Ethics, part of the Open Search Foundation, which is developing a "values compass" in tandem with the technical development.

Findings

All three cases are located within the (supranational) EU, whether the context is national, state, or federated. Their analysis indicates that digital sovereignty can be considered as a personal value, yet sovereignty is not an exclusively individual, ethical value, but a societal and political one. Without a static definition, digital sovereignty then becomes a "discursive claim" as there is no single understanding of the term, which in turn is "crucial for its politics and effects and how it shapes the EU's power" (Adler-Nissen & Eggeling, 2024, p. 3). Nanni, Bizzaro, and Napolitano (2024, 4) declare that "[a]t a high level, digital sovereignty is a process rather than a status." This is echoed by the ZuFiSH and OWI cases, where well-structured problems are iterated at different stages, between knowledge and practice, which include community data as well as community voice and encompasses those of individual participants who take part. This restructuring of a problem, that of defining digital sovereignty, could follow the lead of indigenous communities' knowledge of land and (digital) communality, which promotes well-being for all and reflects how maintenance, sustainability, and stewardship are values that could also be included in the concept of digital sovereignty. Cammaerts and Mansell suggest how policy and regulatory debate should incorporate the views of an "emancipated citizenry" and enable the creation of alternatives (Cammaerts & Mansell, 2020, pp. 147–148). In this manner, digital sovereignty can reflect another form of organization that is neither private nor state but federated and distributed as a form of "digital communality" as with the open web index (Ridgway 2025, forthcoming). By demonstrating how, during the past decade and a half, the terminologies surrounding digital sovereignty have been transformed from traditional nation-state policy-making and individual concerns, the article illuminates how digital infrastructures co-shape individual and collective experiences regarding digital sovereignty from a citizen and user perspective.

1 https://openwebsearch.eu

References

Adler-Nissen, R., & Eggeling, K. A. (2024). The discursive struggle for digital sovereignty: Security, economy, rights and the cloud project Gaia-X. *JCMS: Journal of Common Market Studies, 62*(4), 993–1011. https://doi.org/10.1111/jcms.13594

Calderaro, A., & Blumfelde, S. (2022). Artificial intelligence and EU security: The false promise of digital sovereignty. *European Security, 31*(3), 415–434. https://doi.org/10.1080/09662839.2022.2101885

Cammaerts, B., & Mansell, R. (2020). Digital platform policy and regulation: Toward a radical democratic turn. *International Journal of Communication, 14*. https://ijoc.org/index.php/ijoc/article/view/11182/2901

Couture, S., & Toupin, S. (2019). What does the notion of "sovereignty" mean when referring to the digital? *New Media & Society, 21*(10), 2305–2322. https://doi.org/10.1177/1461444819865984

Floridi, L. (2020). The fight for digital sovereignty: What it is, and why it matters, especially for the EU. *Philosophy & Technology, 33*(3), 369–378. https://doi.org/10.1007/s13347-020-00423-6

Herzog, C., Zetti, D., & Preiß, R. (2024). Digital sovereignty as an ill-structured (or wicked?) problem. Weizenbaum Conference 2024: Uncertain Journeys Into Digital Futures.

Hutchinson, J., Stilinovic, M., & Gray, J. E. (2024). Data sovereignty: The next frontier for internet policy? *Policy & Internet, 16*(1), 6–11. https://doi.org/10.1002/poi3.386

Nanni, R., Bizzaro, P. G., & Napolitano, M. (2024). The false promise of individual digital sovereignty in Europe: Comparing artificial intelligence and data regulations in China and the European Union. *Policy & Internet.* Advance online publication. https://doi.org/10.1002/poi3.424

Paris, B. S., Cath, C., & West, S. M. (2024). Radical infrastructure: Building beyond the failures of past imaginaries for networked communication. *New Media & Society, 26*(11), 6366–6393. https://doi.org/10.1177/14614448231152546

Pohle, J., & Thiel, T. (2020). Digital sovereignty. *Internet Policy Review, 9*(4). https://doi.org/10.14763/2020.4.1532

Pohle, J., & Thiel, T. (2021). Digitale Souveränität: Von der Karriere eines einenden und doch problematischen Konzepts. In C. Piallat (Ed.), *Digitale Gesellschaft* (1st ed., Vol. 36, pp. 319–340). transcript Verlag. https://doi.org/10.14361/9783839456590-014

Ridgway, R. (2025). Designing digital sovereignty – An open federated EU web index for search. *Communication +1*. University of Amherst Press. *(Forthcoming)*

Simon, H. A. (1973). The structure of ill-structured problems. *Artificial Intelligence, 4*(3–4), 181–201. https://doi.org/10.1016/0004-3702(73)90011-8

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

A Relational Approach to Digital Sovereignty

Christian Herzog \\ University of Lübeck \\ Lübeck \\ Germany
christian.herzog@uni-luebeck.de

Jason Branford \\ University of Hamburg \\ Hamburg \\ Germany
jason.branford@uni-luebeck.de

KEYWORDS
digital sovereignty; relational ethics; relational autonomy; digital transformation; public administration

Abstract

The notion of digital sovereignty is subject to several interpretations, ranging from the individual to the territorial and gradations in between. Such shifts in focus are intended to offer conceptual clarity in a domain still finding its footing; however (with few exceptions), comments and critiques of the underlying moral philosophies they suggest remain sorely limited in these discussions. In contrast, this contribution advances an approach to understanding digital sovereignty and its challenges by explicitly building on relational ethics—a framework that is currently gaining interest within ethical reflections of socio-technical domains. Relational approaches acknowledge the increasingly complex socio-technical relations that we find ourselves embedded in. A core strength of relational approaches follows from its emphasis on joining a high regard for the consideration of concrete individual lifeworlds and lived experiences with constructive critiques of structural conditions. We argue that the quest for digital sovereignty requires making this connection, as digital transformations must avoid creating or exacerbating structural injustices, many of which are best understood from individual empirical accounts.

The Digital Sovereignty Discourse

Digital sovereignty has enjoyed various conceptualizations in recent research and public discourse (see, e.g., Pohle and Thiel, 2021; Couture and Toupin, 2019; Floridi, 2020; ZEIT STIFTUNG BUCERIUS,

2024; Calderaro and Blumfelde, 2022). Roberts et al. (2021, p. 6) offer an "overarching definition" by taking digital sovereignty as "a form of legitimate, controlling authority [...] over—in the digital context—data, software, standards, services, and other digital infrastructure". This marks an attempt to make the definition independent from local, temporal, and perspectival conditions, and relies mostly on political power as a concept for legitimacy and control or the (legitimately transferable) ability to influence, e.g., regulations (Floridi 2020, p. 371). In line with Herzog et al. (2024), who consider epistemic iterations between different conceptualizations of digital sovereignty, we acknowledge that the plurality of contributions by different epistemic communities (e.g., scholars, politics, and civil society) not only advances the understanding of digital sovereignty but also supports the identification of practical steps for guiding a just, inclusive, efficient, and effective digital transformation. However, we contend that—with few notable exceptions (e.g., Braun and Hummel, 2024)—digital sovereignty remains underdeveloped from the perspective of moral philosophy. Amending this by drawing on the relatively recent strand of relational ethics (e.g., Birhane, 2021), on the one hand, and relational conceptualizations of autonomy, on the other, promises vital insights that bear on the question of what it is—with respect to digital sovereignty—that societies should try to achieve—and how.

Relational Autonomy and Ethics

Roberts et al. (2021) still emphasize a nation's (as opposed to an individual's) ability to take control over the digital. However, as others have proposed, individual and territorial notions of digital sovereignty are intimately connected (Floridi, 2020). This, of course, also conceives of citizens' capacities to exert influence by voting, petitioning, or standing for candidacy in institutions. Accordingly, a major concern here appears to be mechanisms of political influence—and, hence, political autonomy. There is, as such, a tendency to overlook issues of personal autonomy, which we take as an individual's capacity to authentically express oneself in line with his or her individual values and identity while also considering social relations. It is the social embeddedness of autonomy that defines relational accounts (Mackenzie & Stoljar, 2000), which disavow those overly individualistic accounts of autonomy preeminent in the debate that fail to recognize the role of social relations in value and identity formations (cf., Nagel and Reiner, 2013; Oshana, 2020).

We challenge the view that redefining autonomous agency beyond purely individualistic terms weakens claims to citizen control in digital societies, or, in other words, that endorsing relational autonomy means endorsing paternalism (cf., Humphries, 2025). Rather than negating the ethical significance of self-determination, we contend that in the context of the digital transformation, relational perspectives on autonomy can help identify external factors (relations) that help sustain personal (and political) autonomy and distinguish these from those that undermine it. These external factors also mark possible connections between concrete individual experiences and contingencies within individual lifeworlds and relevant structural conditions established or abolished by the digital transformation. Paired with the epistemic commitments of relational ethics—e.g., by

focusing on marginalized perspectives—this promises a clearer and more nuanced view of what is morally at stake during the digital transformation.

Relational Ethics' Potential for Substantive Digital Sovereignty

We propose to characterize relational ethics' and autonomy's potential to contribute to an understanding of digital sovereignty as a way of realizing *substantive* digital sovereignty—i.e., as a way of enacting specific and socially embedded ways for citizens to exercise their autonomy, both in light of their individual situations and despite specific hardships, but also as socially embedded beings, who may rely on relations, such as within social and political movements, to make their voices heard. Relational ethics directs attention toward marginalized groups, as, e.g., citizens who may rely most on government aid often face additional hardships in an increasingly digital world (cf. Dugdale et al., 2005; D'Ignazio and Klein 2020, p. 105). In turn, acknowledging relational autonomy, specifically, points to empowering individuals to take part in—but also direct or even reject—digitalization agendas not only through educational means or by providing access to digital infrastructures (cf. Herzog and Zetti, 2023) but also by facilitating and recognizing social embeddedness as contributing to value formation. This would explicitly require countering tendencies to regard citizens as individual "customers" of digital public administration services and instead take seriously their capacity to engage in democratic discourse (Pohle and Thiel, 2020; Bekkers and Zouridis, 1999). In practical terms, this calls for commitment toward participatory engagement of citizens by purveyors of the digital transformation, which can also be a viable way to more explicitly consider directing advantages resulting from a renewal of public administration toward those most in need.

References

Birhane, A. (2021) "Algorithmic Injustice: A Relational Ethics Approach." *Patterns* 2(2): 100205. https://doi.org/10.1016/j.patter.2021.100205

Bekkers, V. J. J. M., & Zouridis, S. (1999). Electronic Service Delivery in Public Administration: Some Trends and Issues. *International Review of Administrative Sciences* 65 (2): 183–95. https://doi.org/10.1177/0020852399652004

Braun, M., & Hummel, P. (2024). Is Digital Sovereignty Normatively Desirable? *Information, Communication & Society*, April, 1–14. https://doi.org/10.1080/1369118X.2024.2332624.

Calderaro, A., & Blumfelde, S. (2022). Artificial Intelligence and EU Security: The False Promise of Digital Sovereignty. *European Security* 31(3): 415–34. https://doi.org/10.1080/09662839.2022.2101885

Couture, S., & Toupin, S. (2019). What Does the Notion of 'Sovereignty' Mean When Referring to the Digital? *New Media & Society* 21(10): 2305–22. https://doi.org/10.1177/1461444819865984

D'Ignazio, C., & Klein, L. F. (2020). *Data Feminism*. Strong Ideas. Cambridge: The MIT Press.

Dugdale, A., Daly, A., Papandrea, F., & Maley, M. (2005). Accessing E-Government: Challenges for Citizens and Organizations. *International Review of Administrative Sciences* 71(1): 109–18. https://doi.org/10.1177/0020852305051687

Floridi, L. (2020). The Fight for Digital Sovereignty: What It Is, and Why It Matters, Especially for the EU. *Philosophy & Technology* 33(3): 369–78. https://doi.org/10.1007/s13347-020-00423-6

Herzog, C., & Zetti, D. (2023). Digitally Aided Sovereignty: A Suitable Guide for the E-Government Transformation? *Proceedings of the Weizenbaum Conference 2022: Practicing Sovereignty – Interventions for Open Digital Futures*, edited by Bianca Herlo and Daniel Irrgang, 4–14. Berlin: Weizenbaum Institute for the Networked Society – The German Internet Institute. https://doi.org/10.34669/WI.CP/4.1

Herzog, C., Zetti, D., & Preiß, R. (2024). Digital Sovereignty as an Ill-Structured (or Wicked?) Problem. *Weizenbaum Conference 2024. Uncertain Journeys Into Digital Futures*. Berlin.

Humphries, J. (2025). Defending Relational Autonomy. *Moral Philosophy and Politics*, January. https://doi.org/10.1515/mopp-2023-0053

Mackenzie, C. (with Stoljar, N.) (2000). Relational Autonomy: Feminist Perspectives on Automony, Agency, and the Social Self. Oxford University Press, Incorporated.

Nagel, S. K., & Reiner, P. B. (2013). Autonomy Support to Foster Individuals' Flourishing. *The American Journal of Bioethics* 13(6): 36–37. https://doi.org/10.1080/15265161.2013.781708

Oshana, M. (2020). Relational Autonomy. *The International Encyclopedia of Ethics*, edited by Hugh LaFollette, 1st ed., 1–13. Wiley. https://doi.org/10.1002/9781444367072.wbiee921

Pohle, J., & Thiel, T. (2021). Digitale Souveränität: Von der Karriere eines einenden und doch problematischen Konzepts." *Digitale Gesellschaft*, edited by Chris Piallat, 1st ed., *36*, 319–40. Bielefeld, Germany: transcript Verlag. https://doi.org/10.14361/9783839456590-014

Pohle, J., & Thiel, T. (2020). Digital Sovereignty. *Internet Policy Review 9*(4). https://doi.org/10.14763/2020.4.1532

Roberts, H., Cowls, J., Casolari, F., Morley, F., Taddeo, M., & Floridi, L. (2021). Safeguarding European Values with Digital Sovereignty: An Analysis of Statements and Policies. *Internet Policy Review 10*(3). https://doi.org/10.14763/2021.3.1575

ZEIT STIFTUNG BUCERIUS (2024). “Es geht um unsere Demokratie! Friedensnobelpreisträgerin Ressa und KI-Forscherin Whittaker diskutieren mit uns in Berlin.” ZEIT STIFTUNG BUCERIUS. July 10, 2024. https://www.zeit-stiftung.de/themen/thema/202-es-geht-um-unsere-demokratie-friedensnobelpreistraegerin-ressa-und-ki-forscherin-whittaker-diskutieren-mit-uns-in-berlin

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Gesturing Toward Decolonial ICT

Tech Discernment for Emancipatory Technologies

Dorian Cavé \\ Weizenbaum Institute \\ Berlin \\ Germany
dorian.cave@weizenbaum-institut.de

Rainer Rehak \\ Weizenbaum Institute \\ Berlin \\ Germany
rainer.rehak@weizenbaum-institut.de

KEYWORDS
ICT; decolonial approaches; social change; collective emancipation

Introduction

Considering the coloniality of current digital tools and their massive global social and ecological impacts, we inquire how sociocultural information and communication technology (ICT) assemblages may support the emergence of sustainable and emancipatory 'technodiverse futures and technological pluriverses' (Ricaurte, 2024 p. 5). How could digitalization support emancipatory struggles, be it in the Global North or the Global South, to weave a rhizomatic pluriverse of transformative alternatives aiming to replace the current monocultural and destructive incarnation of digital technologies?

We argue for taking the perspective of grassroots social movements and community initiatives as seeds of change, and for de-centering technologies by focusing on socio-technical praxes. For ICT to play a relevant role in these struggles, transformative initiatives might carry out two tasks: cultivating tech discernment; and establishing emancipatory ICT-enabled sociocultural assemblages. We illustrate our claim using two practical examples and then conclude.

Tech Discernment

To motivate change, such initiatives should first cultivate sophisticated tech discernment with regards to the societal, ecological, and political dimensions of dominant ICT. These tools reinforce and (re)produce authoritarian power structures, and legitimize dominant techno-solutionist and

control-oriented discourses, even in the name of sustainability transitions (Stirling, 2019; Pearson, 2024; Angus, 2022; Rehak, 2024), have devastating impacts on ecosystems and human bodies (Santarius et al., 2023; The Shift Project, 2024; Tréguer and Trouvé, 2017), are designed, produced, and marketed following extractive and (neo)colonial logics including the necessary infrastructures (Brodie, 2023; Ricaurte, 2019; Couldry and Mejias, 2023; Kwet, 2019), and largely realize an idea of "freedom" that is fundamentally individualistic and aims to sever the material, social, and ecological entanglements defining the human condition (Berlan, 2021; Jochum, 2021; Tréguer, 2023) – as opposed to freedom as the collective liberation from structures of domination, which may produce and sustain collective well-being.

Emancipatory ICT-Enabled Sociocultural Assemblages

We argue that informed by such tech discernment, communities and social movements may then adopt and establish emancipatory ICT-enabled sociocultural assemblages, allowing for transformations to take place beyond modern-colonial ways of being.

Such assemblages should notably be, in no particular order of importance (a) critical, in that they explicitly aim at dismantling structures of domination, by adopting decolonial (Alvarado Garcia et al., 2021; Manjarrez, 2023; Guerrero Millan, Nissen, and Pschetz, 2024), feminist (Varon and Egaña Rojas, 2024; Toupin and Hache, 2015), and care- and relationship-centric approaches (Sursiendo, 2019; Emmer et al., 2020; Arora et al., 2020; Zakharova and Jarke, 2024) to digital technologies, (b) degrowth-oriented (Parrique, 2019; Kerschner et al., 2018; Pansera, Ehlers, and Kerschner, 2019), thus aiming to minimise the social and ecological impacts of ICT; and (c) democratic and convivial (Beinsteiner, 2020; Christiaens, 2022; Vetter, 2018; Illich, 1973) by placing a strong emphasis on the political and epistemological implications of the digital tools being used.

Integrating these dimensions, we extend the framework of critiques of modernity proposed by Andreotti et al. (2015), consisting of the soft, radical and beyond reform categories. By doing so, a typology of social-change-oriented digital theories and praxes emerges, hinting towards ICT approaches likely to support the unraveling of modernity-coloniality. We argue that truly decolonial approaches to ICT need to have a sensitivity to the ontology and epistemology that any praxes embody (Machado de Oliveira, 2021; Escobar, 2022).

Two Examples of Community Networks

Examples of community networks deployed in two different contexts may help to illustrate what forms of social change may take place, or fail to emerge, depending on the depth of critique and tech discernment embodied in their ICT praxes.

For instance, democratic and explicitly political processes have been developed in Mexico to enable Indigenous rural communities to develop their own community networks from a critical

perspective, instead of following a hegemonic, 'developmentalist' approach (Baca-Feldman et al., 2018; APC, 2021; Parra Hinojosa and Baca-Feldman, 2021). These processes have led to the adoption and deployment of tools that could be considered as inherently convivial and degrowth-oriented. Such networks can be understood as fitting the beyond-reform category of the typology we suggest above, as they do not seek to support modern-colonial approaches to ICT, but rather 'hack' existing tools for their purposes.

On the other hand, commons-based peer production networks that connect various initiatives in several European countries and beyond (Kostakis and Tsiouris, 2024; Kostakis et al., 2018; Kostakis, Niaros, and Giotitsas, 2023), while degrowth-oriented and convivial, seem to fit the radical reform category, as they are more firmly inscribed within a modernist paradigm.

Conclusion

Making ICT fully sustainable requires fundamental system changes (Hausknost, 2020) reconfiguring all stages of ICT design, production, use, and disposal, as well as the very deployment of the Internet infrastructure, from data center governance to sea cable ownership, following the imperatives of a degrowth agenda challenging incumbent powers and undemocratic institutions preventing ordinary citizens from being political actors. Therefore, any use of ICT in the short term is bound to remain unsustainable, and since they are likely unavoidable as part of struggles for civilizational transformation, they need to be the object of continuous scrutiny.

References

Alvarado Garcia, A., Maestre, J. F., Barcham, M., Iriarte, M., Wong-Villacres, M., Lemus, O. A., Dudani, P., Reynolds-Cuéllar, P., Wang, R., & Cerratto Pargman, T. (2021). Decolonial pathways: Our manifesto for a decolonizing agenda in HCI research and design. In *Extended abstracts of the 2021 CHI conference on human factors in computing systems* (pp. 1–9). ACM. https://doi.org/10.1145/3411763.3450365

Andreotti, V. de O., Stein, S., Ahenakew, C., & Hunt, D. (2015). Mapping interpretations of decolonization in the context of higher education. *Decolonization: Indigeneity, Education & Society, 4*(1). https://jps.library.utoronto.ca/index.php/des/article/view/22168

Angus, S. D. (2022). How liberating is liberation technology? In C. Watkin & O. Davis (Eds.), *New interdisciplinary perspectives on and beyond autonomy* (pp. 101–114). Routledge. https://doi.org/10.4324/9781003331780

APC. (2021). Technological autonomy as a constellation of experiences: A guide to collective creation and development of training programmes for technical community promoters.

Association for Progressive Communications. https://www.apc.org/en/pubs/technological-autonomy-constellation-experiences-guide-collective-creation-and-development

Arora, S., Van Dyck, B., Sharma, D., & Stirling, A. (2020). Control, care, and conviviality in the politics of technology for sustainability. *Sustainability: Science, Practice and Policy, 16*(1), 247–262. https://doi.org/10.1080/15487733.2020.1816687

Baca-Feldman, C. F., Bloom, P., Gómez, M., & Huerta, E. (2018). Community networks in Mexico: A path towards technological autonomy in rural and indigenous communities. In A. Finley (Ed.), *Global information society watch 2018: Community networks* (pp. 178–183). Association for Progressive Communications. https://www.giswatch.org/2018-community-networks

Beinsteiner, A. (2020). Conviviality, the internet, and AI. Ivan Illich, Bernard Stiegler, and the question concerning information-technological self-limitation. *Open Cultural Studies, 4*(1), 131–142. https://doi.org/10.1515/culture-2020-0013

Berlan, A. (2021). Terre et liberté: La quête d'autonomie contre le fantasme de délivrance. Éditions La Lenteur.

Brodie, P. (2023). Data infrastructure studies on an unequal planet. *Big Data & Society, 10*(1), 20539517231182402. https://doi.org/10.1177/20539517231182402

Christiaens, T. (2022). Convivial autonomy in platform capitalism. In O. Davis & C. Watkin (Eds.), *New interdisciplinary perspectives on and beyond autonomy* (pp. 69–82). Routledge. https://doi.org/10.4324/9781003331780-7

Couldry, N., & Mejias, U. A. (2023). The decolonial turn in data and technology research: What is at stake and where is it heading? *Information, Communication & Society, 26*(4), 786–802. https://doi.org/10.1080/1369118X.2021.1986102

Emmer, P., Salas Neves, B., Rivas, C., & Schweidler, C. (2020). *Technologies for liberation: Toward abolitionist futures*. Astraea Lesbian Foundation for Justice and Research Action Design. https://astraeafoundation.org/wp-content/uploads/2020/12/Technologies-for-Liberation-1.pdf

Escobar, A. (2022). Reframing civilization(s): From critique to transitions. *ARQ (Santiago), (111)*, 24–41. https://doi.org/10.4067/S0717-69962022000200024

Guerrero Millan, C., Nissen, B., & Pschetz, L. (2024). Cosmovision of data: An Indigenous approach to technologies for self-determination. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). ACM. https://doi.org/10.1145/3613904.3642598

Hausknost, D. (2020). The environmental state and the glass ceiling of transformation. *Environmental Politics, 29*(1), 17–37. https://doi.org/10.1080/09644016.2019.1680062

Illich, I. (1973). *Tools for conviviality*. Harper & Row.

Jochum, G. (2021). Dialectics of technical emancipation—Considerations on a reflexive, sustainable technology development. *NanoEthics, 15*(1), 29–41. https://doi.org/10.1007/s11569-021-00387-7

Kerschner, C., Wächter, P., Nierling, L., & Ehlers, M.-H. (2018). Degrowth and technology: Towards feasible, viable, appropriate and convivial imaginaries. *Journal of Cleaner Production, 197*, 1619–1636. https://doi.org/10.1016/j.jclepro.2018.07.147

Kostakis, V., Latoufis, K., Liarokapis, M., & Bauwens, M. (2018). The convergence of digital commons with local manufacturing from a degrowth perspective: Two illustrative cases. *Journal of Cleaner Production, 197*, 1684–1693. https://doi.org/10.1016/j.jclepro.2016.09.077

Kostakis, V., Niaros, V., & Giotitsas, C. (2023). Beyond global versus local: Illuminating a cosmolocal framework for convivial technology development. *Sustainability Science, 18*(5), 2309–2322. https://doi.org/10.1007/s11625-023-01378-1

Kostakis, V., & Tsiouris, N. (2024). How to unite local initiatives for a more sustainable global future. *Sustainable Futures, 7*, 100187. https://doi.org/10.1016/j.sftr.2024.100187

Kwet, M. (2019). Digital colonialism: South Africa's education transformation in the shadow of Silicon Valley [PhD thesis, Rhodes University]. https://papers.ssrn.com/abstract=3496049

Machado de Oliveira, V. (2021). Hospicing modernity: Facing humanity's wrongs and the implications for social activism. North Atlantic Books.

Manjarrez, L. E. (2023). Towards pluriversal views of digital technologies: The experiences of community and Indigenous radios in Chiapas, Mexico. *Tapuya: Latin American Science, Technology and Society, 6*(1), 2254629. https://doi.org/10.1080/25729861.2023.2254629

Pansera, M., Ehlers, M.-H., & Kerschner, C. (2019). Unlocking wise digital techno-futures: Contributions from the degrowth community. *Futures, 114*, 102474. https://doi.org/10.1016/j.futures.2019.102474

Parra Hinojosa, D., & Baca-Feldman, C. F. (2021). Let us rethink communication technologies: Methodological proposals to design and implement community communication projects. Redes por la Diversidad, Equidad y Sustentabilidad A.C. https://ed8169c2-0818-439d-b473-11f6b06914e9.filesusr.com/ugd/68af39_4bda81212a9c42c4b15f43c9a1e208f7.pdf

Parrique, T. (2019). *The political economy of degrowth* [PhD thesis, Université Clermont Auvergne; Stockholms universitet]. https://theses.hal.science/tel-02499463v1

Pearson, J. S. (2024). Defining digital authoritarianism. *Philosophy & Technology, 37*(2), 73. https://doi.org/10.1007/s13347-024-00754-8

Ricaurte, P. (2019). Data epistemologies, the coloniality of power, and resistance. *Television & New Media, 20*(4), 350–365. https://doi.org/10.1177/1527476419831640

Ricaurte, P. (2024). The digitalization imperative. *Dialogues on Digital Society*, 29768640241262263. https://doi.org/10.1177/29768640241262263

Rehak, R. (2024). On the (im)possibility of sustainable artificial intelligence. In T. Züger & H. Asghari (Eds.), *AI systems for the public interest*. *Internet Policy Review, 13*(3). https://doi.org/10.5281/zenodo.14283597

Santarius, T., Bieser, J. C. T., Frick, V., Höjer, M., Gossen, M., Hilty, L. M., Kern, E., Pohl, J., Rohde, F., & Lange, S. (2023). Digital sufficiency: Conceptual considerations for ICTs on a finite planet. *Annals of Telecommunications, 78*(5), 277–295. https://doi.org/10.1007/s12243-022-00914-x

Stirling, A. (2019). Sustainability and the politics of transformations: From control to care in moving beyond modernity. In J. Meadowcroft et al. (Eds.), *What next for sustainable development?: Our common future at thirty* (pp. 219–238). Edward Elgar Publishing. https://www.elgaronline.com/edcollchap/edcoll/9781788975193/9781788975193.00008.xml

Sursiendo. (2019, July 16). Encuentro hackfeminista en Chiapas: Estar en analógico para construir entornos digitales más dignos para nosotras. *Sursiendo*. https://sursiendo.org/2019/07/encuentro-hackfeminista-en-chiapas-estar-en-analogico-para-construir-entornos-digitales-mas-dignos-para-nosotras/

The Shift Project. (2024, March 29). "Virtual worlds and networks facing the dual carbon constraints" – New reports release – The Shift Project. *The Shift Project*. https://theshiftproject.org/en/article/virtual-worlds-and-networks-new-reports-release/

Toupin, S., & Hache, A. (2015). Feminist autonomous infrastructure. In A. Finlay (Ed.), *Global information society watch 2015: Sexual rights and the internet* (pp. 22–25). Association for Progressive Communications and Hivos.

Tréguer, F. (2023). Contre-histoire d'Internet: Du XVe siècle à nos jours. Agone.

Tréguer, F., & Trouvé, G. (2017, May 27). Le coût écologique d'internet est trop lourd, il faut penser un internet low-tech. *Reporterre, le média de l'écologie*. https://reporterre.net/Le-cout-ecologique-d-internet-est-trop-lourd-il-faut-penser-un-internet-low

Varon, J., & Egaña Rojas, L. (2024). *Compost engineers and sus saberes lentos: A manifest for regenerative technologies* (J. Varon, Trans.). Coding Rights. http://codingrights.org/docs/compost_engineers

Vetter, A. (2018). The matrix of convivial technology – Assessing technologies for degrowth. *Journal of Cleaner Production, 197*, 1778–1786. https://doi.org/10.1016/j.jclepro.2017.02.195

Zakharova, I., & Jarke, J. (2024). Care-ful data studies: Or, what do we see, when we look at datafied societies through the lens of care? *Information, Communication & Society, 27*(4), 651–664. https://doi.org/10.1080/1369118X.2024.2316758

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Algorithmic Domination

Konstanze Möller-Jansen \\ TU Dresden \\ Dresden \\ Germany
konstanze.moeller-jansen@tu-dresden.de

KEYWORDS
domination, power, freedom as non-domination, big tech, algorithmic regulation, algorithmic domination

Introduction

Algorithmic decision systems are increasingly involved in mediating social relationships (Bucher, 2018; Elkin-Koren & Perel, 2018; Lazar, forthcoming). They help us to filter through the vast array of options on platforms and search engines, help us choose to select the highest-rated restaurant or hotel, or find the fastest route to the nearest supermarket. They ensure that we mostly see content we like and ads that match our algorithmically determined interests. One reason for the targeted application of these complex, socio-technical systems is that they are useful tools for regulating the behaviour of larger groups of people by steering it into desired patterns of behaviour, called "algorithmic regulation" (Brownsword, 2005; Katzenbach & Ulbricht, 2019; König, 2020; Yeung, 2018).

The widespread use of algorithmic decision systems in many commercial contexts has consequently raised concerns over skewed and problematic power relations. The general concern seems to be that we are faced with unequal power relations that disadvantage users against opaque technological systems, or against powerful agents that use them, such as large internet platforms and technology companies (Crawford, 2021; Liu, 2018; Miceli et al., 2021; Susskind, 2022).

In light of the current literature on what problematic about this situation, I agree that this concern is best framed as a problem of power: the use of these artefacts not only enhances the ability of certain agents to achieve their own goals and ends, but also enables them to influence and constrain others in achieving theirs. Consequently, we are faced with two problems: the concentration of power in the hands of certain agents, such as internet platforms and tech companies, and the limited agency on the part of those subject to that power. For many, the latter problem consists of agents being subject to data surveillance, to manipulation on internet platforms, or opaque algorithmic decision systems that determine and frame their available options. In short, some fear that algorithmic decision systems undermine the freedom of those exposed to them. The question now is how to bridge the gap between these two issues of power and freedom.

In my view, the link between power and freedom is 'domination' (Lovett, 2010, 2022; Pettit, 1997, 2012). According to this concept, to be unfree is to be dependent on and subject to another agent's uncontrolled and arbitrary social power. I will argue that this concept offers a fitting lens to explain the problem in the context described above. It is not that 'Big Tech'1 controls our decisions or coerces us into certain behaviours. Rather, the main normative problem is that such companies dominate end-users, other economic actors, and even governments (to varying degrees) in virtue of two types of power: 'impositional power' (McCammon, 2015) and 'systemic power' (Gädeke, 2020; Jugov, 2024).

First, they are able to unilaterally determine the basic conditions of any interaction or communication mediated via their platforms, services, and products. Their ability impose conditions of use and access for basic utilities constitutes a form of 'impositional power'. This is the ability of an agent to determine the costs of cooperation in a relationship of strategic dependence. Furthermore, based on their ability to engage in processes of surveillance and prediction as social sorting (Lyon, 2003), and the pervasive use of algorithmic regulation they can a) generate opaque algorithmic categorizations and b) influence people's habits and preferences. By subjecting users to these novel systems of classification and algorithmic behavioral management, Big Tech thereby has the ability to influence the social rules and practices that organise how those dependent on them interact and communicate with each other. This constitutes a form of arbitrary 'systemic power'.

The Outsized Influence of Big Tech

As an unstructured group, Big Tech possesses control over goods and services that have proven to be fundamental to the functioning of our information society. This includes control over "advertising networks, login services, cloud hosting, app stores, payment systems, data analytics, video hosting, geospatial and navigation services, search functionalities, operating systems, and more recently, artificial intelligence (AI) services" (Poell et al., 2018). Their advantageous and outstanding market position and influence are built, most notably, on two pillars: platform power and control over data. This creates a situation, in which end-users, competitors and other economic actors, and even governments are dependent on the basic infrastructure offered and controlled by these "new utilities" (Rahman, 2018).

We can characterise their influence further by referring to their ability to engage in surveillance and prediction as a process of social sorting. Data is not merely collected; it is actively generated. Users are compelled to share and produce data whenever and wherever they engage in interaction and communication mediated by a large platform. Data is generated with the specific aim of making algorithmic predictions in order to categorise individuals based on the risk they pose of deviating from a descriptive norm or standard. This is known as "algorithmic regulation" (Yeung, 2018; Yeung

1 'Big Tech' is a commonly used term for a group of internet platforms and technology companies, mainly Google and Alphabet, Meta and Facebook, Amazon, and Microsoft. Which specific companies belongs to this group, or which criterion determines 'group membership', is not relevant for our current purposes.

& Lodge, 2019), whereby the behaviour of populations of users is steered towards desired patterns of behaviour through a constant reconfiguration of a person's choice architecture.

What Is (Normatively) Wrong About This Kind of Influence?

To characterize what is normatively problematic about the kind of power that we see here, we can draw on the concept of 'domination'. In our context, domination is based on two types of arbitrary power: 'impositional power' and 'systemic power'.

First of all, I argue that Big Tech's control over basic infrastructural goods and services has created a situation of asymmetric dependency of stakeholders such as end-users, other economic actors and even governments. Domination based on arbitrary impositional power arises, when agent B depends on another agent A in a way that A does not depend on B, and A is able to make B's costs of non-cooperation higher than those of cooperation. Big Tech unilaterally imposes use and access conditions for increasingly basic utilities, thus determining the costs of cooperation for those dependent on its goods and services. I argue that this largely uncontrolled and arbitrary form of power is a form of algorithmic domination.

Second of all, I would argue that Big Tech can shape people's option sets through a continuously updated, personalized and ever-present reconfiguration of the choice architecture. In this way, they influence the formation of preferences and habits, for example, by nudging users to engage in certain behaviors or to desire things that the Big Tech ecosystem can provide. What sets apart their power is the ability to generate decision rules in the first place: We are exposed to opaque categorisations and classifications produced by their big data-fuelled algorithmic decision systems. This is a form of arbitrary systemic, or 'meta-', power: power over social rules and practices based on the ability to establish a system of relevance through algorithmic categorisation and predictive analytics.

In conclusion, the power of Big Tech is not necessarily one of direct interference, coercion or the imposition of constraints. Rather, it can be understood as a form of domination. This is based on its outsized, unilateral ability to determine the costs of cooperation in an unequal bargaining scheme – in other words, arbitrary impositional power. It is also power over social categorisations, classifications and social rules that influence users' habits and practices – in other words, systemic meta power.

Funding Information and Acknowledgements

Funded by the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) as part of Germany's Excellence Strategy – EXC 2050/1 – Project ID 390696704 – Cluster of Excellence "Centre for Tactile Internet with Human-in-the-Loop" (CeTI) of Technische Universität Dresden.

References

Brownsword, R. (2005). Code, control, and choice: Why East is East and West is West. *Legal Studies, 25*(1).

Bucher, T. (2018). *If...then: Algorithmic power and politics*. Oxford University Press.

Crawford, K. (2021). Atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press.

Elkin-Koren, N., & Perel, M. (2018). Algorithmic governance by online intermediaries. In *Oxford handbook of international economic governance and market regulation* (p. 25). Oxford University Press.

Gädeke, D. (2020). Does a mugger dominate? Episodic power and the structural dimension of domination. *Journal of Political Philosophy, 28*(2), 199–221. https://doi.org/10.1111/jopp.12202

Jugov, T. (2024). Geltungsgründe globaler Gerechtigkeit. Campus Verlag.

Katzenbach, C., & Ulbricht, L. (2019). Algorithmic governance. *Internet Policy Review, 8*(4). https://doi.org/10.14763/2019.4.1424

König, P. D. (2020). Dissecting the algorithmic leviathan: On the socio-political anatomy of algorithmic governance. *Philosophy & Technology, 33*(3), 467–485. https://doi.org/10.1007/s13347-019-00363-w

Lazar, S. (forthcoming). Connected by code: How AI structures, and governs, the ways we relate. Oxford University Press.

Liu, H.-Y. (2018). The power structure of artificial intelligence. *Law, Innovation and Technology, 10*(2), 197–229. https://doi.org/10.1080/17579961.2018.1527480

Lovett, F. (2010). *A general theory of domination and justice*. Oxford University Press.

Lovett, F. (2022). *The well-ordered republic*. Oxford University Press. https://doi.org/10.1093/oso/9780192859556.001.0001

Lyon, D. (Ed.). (2003). Surveillance as social sorting: Privacy, risk, and digital discrimination. Routledge.

McCammon, C. (2015). Domination: A rethinking. *Ethics, 125*(4), 1028–1052. https://doi.org/10.1086/680906

Miceli, M., Posada, J., & Yang, T. (2021). Studying up machine learning data: Why talk about bias when we mean power? (arXiv:2109.08131). *arXiv*. http://arxiv.org/abs/2109.08131

Pettit, P. (1997). *Republicanism: A theory of freedom and government*. Clarendon Press; Oxford University Press.

Pettit, P. (2012). On the people's terms: A republican theory and model of democracy. Cambridge University Press.

Poell, T., Nieborg, D., & Van Dijck, J. (2018, October 10). Platform power and public value. Paper presented at AoIR 2018: The 19th Annual Conference of the Association of Internet Researchers, Montréal, Canada.

Rahman, K. S. (2018). The new utilities: Private power, social infrastructure, and the revival of the public utility concept. *Cardozo Law Review, 39*(5), 1621–1692.

Susskind, J. (2022). *The digital republic: On freedom and democracy in the 21st century* (First Pegasus Books cloth edition). Pegasus Books.

Yeung, K. (2018). Algorithmic regulation: A critical interrogation. *Regulation & Governance, 12*(4), 505–523. https://doi.org/10.1111/rego.12158

Yeung, K., & Lodge, M. (2019). Algorithmic regulation: An introduction. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3483693

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Normative Power and Autonomy in the Digital Era

Juliane Mendelsohn \\ TU Ilmenau \\ Ilmenau \\ Germany
juliane.mendelsohn@tu-ilmenau.de

KEYWORDS
autonomy; power; rule-setting; platforms

Abstract

This theoretical (philosophy and legal theory) paper looks at the impact of the design and normative power of digital platforms and ecosystems on the self-governance of societies and on individual autonomy. One of the more puzzling features of the economic power of digital platforms or ecosystems is not only their reach beyond the parameters of traditional markets (Mendelsohn 2023) but also into the social, political, private and moral spheres of people's existence in society. In this regard, the study of their normative or regulatory power has emerged as a new phenomenon of inquiry for scholars of law and economics and law of the political economy (e.g. Crémer, de Montiye and Schweitzer 2019, 6, 16; Yeung 2026; Delacroix 2020; Gerbrandy and Phoa 2022).

In a narrow sense, the Cremér, Schweitzer and de Montjoye Report on the digital economy for the EU Commission already highlighted that 'marketplace platforms play a regulatory role as they determine the rules of the marketplace and possibly also the rules based on which their clients interact' (Crémer, de Montiye and Schweitzer 2019, 54). Normative power has most broadly been defined as „(the) power to shape current and future normality, normativity, markets, social relations and behaviour, and truth" (Gerbrandy 2022). Normative power thus includes legal or contractual rules, but also design choices and algorithmic patterning and even the creation of new habits (as a source of law). This normative power encroaches not only on the regulatory power of governments, but also on the central ideas of a self-governing private law society (Privatrechtsgesellschaft – Mestmäcker, 2019) and the concepts of individual autonomy and sovereinty.

The paper explores various dimensions of normative power as distinct from other forms of market failure. It looks at how single, powerful actors unilaterally define the correct rules and conditions for large digital spaces and entire sectors of the economy. Sources include contractual conditions, algorithmic design choices and their influence on discourse and law enforcement. In addition to

their power over truth and information, which is the subject of several studies on mis-information, this paper also focusses on the (subtler) private-law question of how such actors not shape and define habits (Delacoix 2020) and influence importance allocative choices with regard to data, digital rights, time and attention. Owing to its enlightenment origins, this paper finally explores ways in which true autonomy can be reclaimed (Pauer-Studer, 2000).

References

Crémer, J., de Montjoye, Y.-A., & Schweitzer, H. (2019). *Competition policy for the digital era* (EU Commission, April 2019). https://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf

Delacroix, S. (2020). Preserving us from regulatory power? Legal normativity and the possibility of agency. In C. Bezemek, M. Potacs, & A. Somek (Eds.), *Vienna lectures on legal philosophy, Vol. 2: Normativism and anti-normativism in legal theory* (pp. 233–250). Mohr Siebeck.

Gerbrandy, A. (2018). *Competition law submission, 2018, S. 1(2).* https://ssrn.com/abstract=3275235

Gerbrandy, A., & Phoa, P. (2022). The power of big tech corporations as modern bigness and a vocabulary for shaping competition law as counter-power. In M. Bennett, H. Brouwer, & R. Claasen (Eds.), *Wealth and power* (p. 166). Routledge.

Mendelsohn, J. (2023). Reconsidering conglomerates – How are digital conglomerates different from those in the past? Theory and implications. *Competition Law Review, 15*(1), 83–103.

Mestmäcker, E. J. (2019). Wettbewerb in der Privatrechtsgesellschaft. Mohr Siebeck.

Pauer-Studer, H. (2020). Autonom leben: Reflexionen über Freiheit und Gleichheit. Suhrkamp Verlag.

Yeung, K. (2016). 'Hypernudge': Big data as a mode of regulation by design. *Information, Communication & Society, 19*(1). *(Originally published as TLI Think! Paper 28/2016)* https://doi.org/10.1080/1369118X.2016.1186713

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Exploration of Mass Comment Campaigns in European Public Consultations Using an LLM

Quentin Bukold \\ Weizenbaum Institute \\ Berlin \\ Germany
quentin.bukold@weizenbaum-institut.de

KEYWORDS
e-participation; public consultation; European Commission; interest group strategy; large language model

Abstract

Mass Comment Campaigns (MCCs) are coordinated efforts by organizations to mobilize their members and supporters to submit near-identical comments. While common in European Public Consultations, they remain understudied. This paper introduces a novel pipeline to identify and analyze MCCs while also introducing a dataset of 714 legislative initiatives (2017–2024) and their feedback entries. An LLM enables cross-linguistic insights into the campaigns content. Applying the pipeline to three high-engagement consultations in the field of agricultural and environmental policy revealed MCCs of varying professionalization and transnational reach. Findings highlight MCCs' dominance in some consultations and their potential role in shaping EU policymaking. This study contributes to understanding European MCCs and calls for further research on their impact and classification.

Introduction

This paper aims to introduce the phenomenon of Mass Comment Campaigns (MCC), which have often been documented in European consultations (Nørbech 2024; Marxsen 2015; Lironi and Peta 2017), yet in-depth analysis has primarily focused on the US context (Balla et al. 2019; 2022). MCCs are coordinated efforts by organizations to mobilize their members and supporters to submit near-identical comments, aiming to influence the rulemaking process by exerting pressure on politicians and civil servants (Balla et al. 2019, 461–76). Public consultations provide an online space for

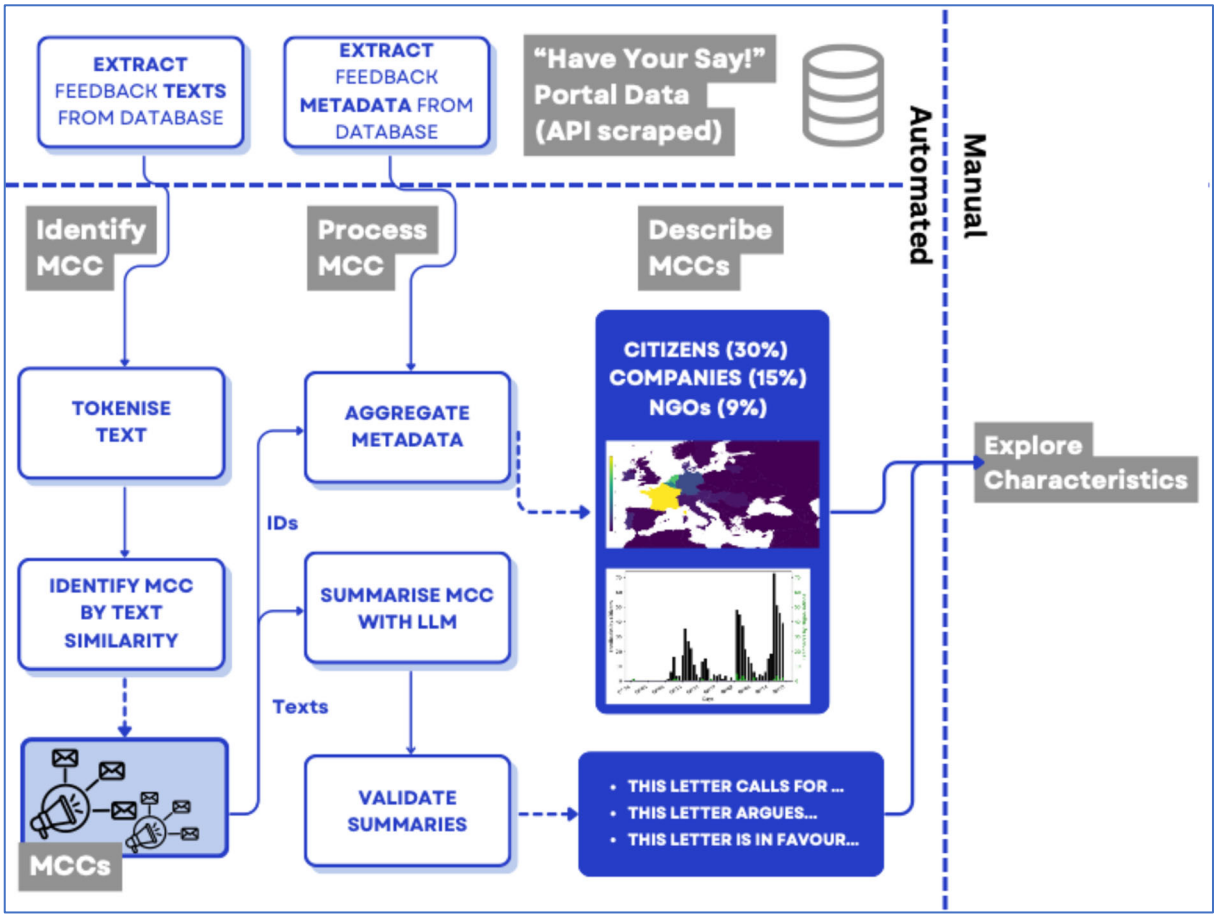
individuals and organizations to submit comments during the development of new proposals for directives and regulations by the Commission. The platform for these consultations features input from major stakeholders such as BlackRock, NABU, and the German BMDV, alongside hundreds of thousands of comments from citizens that constitute a key source of policy information for the Commission (Bunea 2019). Yet, for this integral part of European rulemaking (Garben and Govaere 2018, 217) we lack the methods for identifying and describing the multilingual and in parts extremely voluminous MCCs. This study intends to present a pipeline that allows us to identify and describe MCCs. It also provides insights into a sample of MCCs, exploring novel characteristics of European MCCs. The pipeline is aimed at researchers as a way to better understand the effect of MCCs on European legislation and citizen participation.

Methods in Brief

To lay the foundations for further research into European Public Consultations, 714 legislative initiatives between February 2017 and March 2024 have been scraped, which resulted in around 500,000 open text feedbacks from organizations and individuals. Inspired by Matter's (2020) concept of big public data, the data was collected via the open API endpoints of the consultation platform. The metadata includes the names of all submitting organizations, their transparency register number, type and more.

Potential MCCs were identified by deploying an algorithm which compares tri-grams between feedback texts and computes a similarity score based on the Jaccard Score (Jaccard 1902, 72). This approach enables the clustering of feedback submissions that share highly similar wording while still accounting for minor modifications and extensions made to the pre-formulated texts provided by MCC sponsors. To describe the content of identified MCCs, a large language model (LLM) was used to summarize individual feedback entries (see Fig. 1). These summaries were then validated using Sentence-BERT to prevent hallucination. Extending previous work, this study leverages the LLM Mistral-Nemo for summarizing and translating consultation feedback, enabling cross-linguistic insights (Salas-Girones et al., 2024).

Figure 1: Process of analysis after scraping



Exemplary Analysis and Outlook

The proposed analytical pipeline was applied to three high-engagement initiatives and their public consultations – all of them lying in the policy field of agricultural and environmental policy, which allowed for good comparability. 49 clusters were identified and described. By comparing the translated and summarized content and the submitting organizations, some clusters were grouped further to account for multilingual MCCs or incorrect splitting of some clusters – leading to a total of 10 MCCs.

One MCC by the NGO "Gen-ethisches Netzwerk e.V." and business association "Demeter" was responsible for more than 90 % of the about 70,000 citizen feedbacks submitted during a feedback period concerning new genomic techniques for agricultural use. Further investigation showed the use of a website widget which allowed to submit feedback in multiple languages with minimal effort bringing a kind of low-effort participation into Public Consultations. Another MCC was linked to the 'Save Bees and Farmers' movement. This became clear from the summaries of the content and the presence of the movement's organizations among the contributors of multiple clusters. The MCC was organized for four different languages, spanned over two months and mobilized citizens as well

as companies, academic institutions and consumer organizations. Another MCC, consisting of only 18 feedbacks, had been started and supported by multiple Polish private foresters and state foresters.

The analysis shows that European MCCs come in various levels of professionalization and transnationalization. They dominate some consultations in terms of volume and partly even constitute a channel for state organizations to be heard by the Commission. This paper wants to support the call for more research into MCCs beyond the US context (Balla et al. 2019, 477) and invite you to analyze data on MCCs. Also, research still lacks a unified threshold for what makes comment campaigns Mass Comment Campaigns (Balla et al. 2019, 464). Do we understand them as such when a campaign publicly calls for the submission of texts, i.e. is a 'campaign for the masses', or can a threshold be defined in relation to the size of the political endeavor? Also, what will happen if submitters employ LLMs to diversify their submissions? Will we still be able to tell individual from preformulated feedback? In my presentation, I will provide an overview about my results and present systematic differences between the campaigns.

Funding Information and Acknowledgments

This research was not funded. The data on which this work is based was collected as part of my graduation thesis at the University of Hildesheim. I would like to thank Prof. Dr. Ulrich Heid, Dr. Jessica Schwarz, Dr. Jakob Ohme, Lion Wedel, Anna-Theresa Mayer and Tabea for their support in the preparation of this paper and the work on which it is based.

References

Balla, S. J., Beck, A. R., Cubbison, W. C., & Prasad, A. (2019). Where's the spam? Interest groups and mass comment campaigns in agency rulemaking. *Policy & Internet, 11*(4), 460–479. https://doi.org/10.1002/poi3.224

Balla, S. J., Beck, A. R., Meehan, E., & Prasad, A. (2022). Lost in the flood? Agency responsiveness to mass comment campaigns in administrative rulemaking. *Regulation & Governance, 16*(1), 293–308. https://doi.org/10.1111/rego.12318

Bunea, A. (2019). Stakeholder consultations. In *The Palgrave encyclopedia of interest groups, lobbying and public affairs* (pp. 1–7). Springer International Publishing.

Garben, S., & Govaere, I. (Eds.). (2018). *The EU better regulation agenda: A critical assessment* (Modern Studies in European Law, Vol. 87). Hart.

Jaccard, P. (1902). Lois de distribution florale dans la zone alpine. *Bulletin de la Société Vaudoise des Sciences Naturelles, 38*(144), 69. https://doi.org/10.5169/seals-266762

Lironi, E., & Peta, D. (2017). EU public consultations in the digital age: Enhancing the role of the EESC and civil society organisations. Brussels: EESC. https://doi.org/10.2864/464177

Marxsen, C. (2015). Open stakeholder consultations at the European level—Voice of the citizens? *European Law Journal, 21*(2), 257–280. https://doi.org/10.1111/eulj.12084

Nørbech, I. (2024). Does policy context matter for citizen engagement in policymaking? Evidence from the European Commission's public consultation regime. *European Union Politics, 25*(1), 130–150. https://doi.org/10.1177/14651165231208995

Salas-Girones, E., Murukannaiah, P., & Ingrams, A. (2024). Summarizing public comments on policy proposals using large language models.

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Classifying Informative Short Vertical Videos

Moving Away from External Definitions to a User-Centered Approach

Lion Wedel \\ Weizenbaum Institute \\ Berlin \\ Germany
lion.wedel@weizenbaum-institut.de

KEYWORDS
information perception; user-centered; news use; information use; digital trace data; audience research

Introduction

The rise of short-form video platforms like TikTok has fundamentally altered information dissemination, particularly among younger audiences. A key challenge for communication science researchers is developing scalable classification methods for online news content to understand these changes beyond qualitative analysis. Traditional classification approaches, relying on external categorizations like journalistic outlet lists, are insufficient. They fail to recognize that professional media outlets and journalists are no longer the sole publishers of news, nor do they exclusively publish news content either (Harff and Schmuck 2024; Negreira-Rey, Vázquez-Herrero, and López-García 2022). Furthermore, news and information consumption research faces an evolving understanding of "news" among young adults (Swart and Broersma 2023; Cotter and Thorson 2022). This has prompted researchers to adopt a broader "information" terminology, encompassing anything perceived as new and/or useful (e.g., Kümpel, Anter, and Unkel 2022), thereby including content and actor types beyond traditional journalistic norms.

This shift leaves current research on audience perceptions of news and information lacking the computational tools to utilize fine-grained, user-level digital trace data effectively. Such data offers the potential for unprecedented insights into news dissemination from a user perspective (Ohme et al. 2023). This work addresses this gap. I present and test a novel set of survey items, derived from prior qualitative research, designed to identify key factors driving user perceptions of informative posts and to identify distinct user types. I further introduce a user-centered classifier, incorporating user information perception types, and compare its performance with traditional classification methods.

Methodology

A survey was conducted with a German quota sample (age, gender, education) of short-form video platform users (TikTok, Instagram Reels, YouTube) (N = 2153). Beyond sociodemographic information, participants responded to 44 items, based on previous qualitative research, measuring the influence of various aspects on their perception of informative posts. These items spanned three dimensions: content, actor, and algorithmic selection. Factor analysis was employed to derive eight factors from these items, which then served as the basis for a K-means cluster analysis of participants.

Subsequently, I will train a classifier for each identified user cluster. Leveraging the capabilities of agentic LLMs, we will fine-tune a model for each cluster, using the cluster characteristics as input, thus minimizing the need for extensive training data. I will then compare the performance of our user-centered model(s) with two traditional classification approaches: actor-based and hashtag-based classification.

Preliminary Results and Outlook

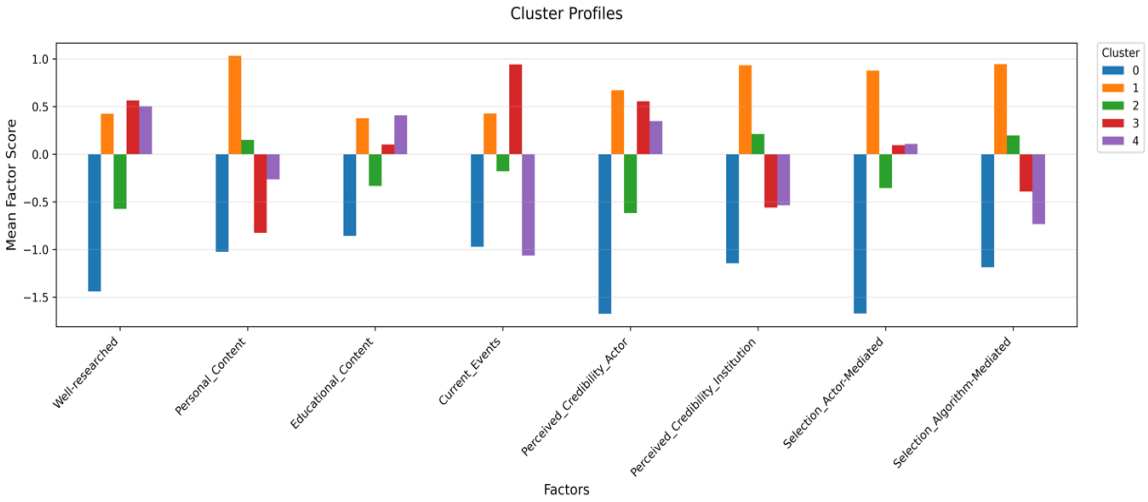
The survey results with the subsequent factor analysis resulted in eight factors across the dimensions: content, actor, and selection type (see Table 1).

Table 1: Resulting factors and their descriptions. The factor analysis was conducted independently for the subset of items of each dimension.

Factor	Dimension	Description
1	Content	Well-researched content (sources etc.)
2	Content	Casual content (friends, trends, etc.)
3	Content	Educative and helpful content
4	Content	Societal relevant content (similar hard news)
5	Actor	Perceived credibility/trustworthiness of the individual actor
6	Actor	Perceived credibility/trustworthiness of the actor's institution/profession
7	Selection	Actor-mediated selection
8	Selection	Algorithmic-mediated selection

Based on these factors, the resulting user clusters reveal distinct user groups, each with a unique combination of factors influencing their perception of online post informativeness (see Figure 1).

Figure 1: Mean factor score across factors for each cluster. (The scores represent the mean deviation of one cluster's users compared to the mean of the overall sample.)



Preliminary analysis suggests interesting differences between clusters. Cluster 0 (blue) scores below average on all factors, while Cluster 1 (orange) scores above average. This may indicate a generally pessimistic view of online information in Cluster 0 and a generally positive view in Cluster 1. The below-average score for perceived actor credibility in Cluster 2 (green) is noteworthy, potentially suggesting a lower appreciation for non-professional news sources like influencers. Clusters 3 and 4 share similar profiles for actor and selection dimensions but differ in their content preferences. Cluster 3 users score highly on traditional news content types (current events) and characteristics (well-researched), while Cluster 4 users score above average on educational content. Further analysis of additional survey data will provide deeper insights into these user types.

These preliminary findings demonstrate the potential of a user-centered approach to understanding information perception on short-form video platforms. I can move beyond traditional content classification methods by identifying distinct user types based on their information evaluation criteria. The next step involves training and evaluating the performance of a classifier, comparing it against actor-based and hashtag-based classification benchmarks. I anticipate that incorporating user perception profiles will enable a more nuanced understanding of information dissemination dynamics. This user-centric approach promises valuable insights for researchers and professionals alike. Crucially, our approach allows us to understand users' news and information consumption not from assumed criteria of informativeness, but from their perspectives.

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

How Citizens Search for Information About Climate Change

Role of Search Suggestions, Political Beliefs, and Intuition

Victoria Vziatysheva \\ University of Bern \\ Bern \\ Switzerland
victoria.vziatysheva@unibe.ch

Mykola Makhortykh \\ University of Bern \\ Bern \\ Switzerland
mykola.makhortykh@unibe.ch

Maryna Sydorova \\ University of Bern \\ Bern \\ Switzerland
maryna.sydorova@unibe.ch

Vihang Jumle \\ University of Bern \\ Bern \\ Switzerland
vihang.jumle@unibe.ch

KEYWORDS
search engines; information-seeking behavior; selective exposure; climate change; search queries; voting

Abstract

In a highly saturated media environment, search engines serve as an important gateway to political information. However, there are still gaps in understanding how citizens interact with these platforms. While numerous studies explore how users navigate search results, less research has been done on how they formulate search queries to find information on various political issues. Yet, the choice of a query is crucial, as it largely determines what information users are exposed to on search engines, which has direct implications for democratic decision-making.

Earlier research on political information-seeking behavior has provided extensive evidence that people tend to favor information confirming their beliefs, a phenomenon also known as selective exposure. Selective exposure to information can pose risks for political decision-making, as it has the potential to limit the informedness of citizens and amplify societal polarization. However, while there is fairly consistent evidence that users select information aligned with their beliefs (e.g.,

Ekström et al., 2022; Robertson et al., 2023), the evidence on selective exposure in what information they initially search for remains fragmented. While some studies find that search query formulation is motivated by personal attitudes and beliefs (e.g., Ekström et al. 2024), others show little to no support for this argument (e.g., van Hoof et al. 2024; Vziatysheva et al. 2024). These discrepancies can be explained by differences in both the topics of search queries (e.g., highly polarizing or less polarizing) and the mode of data collection (e.g., whether respondents are asked to select a query from the curated list of answers or compose it independently).

To better understand the reasons for selective exposure through search engines, we use a representative survey of Swiss citizens (N = 1,070), which explores how voters search for information about an environment-related popular initiative that was voted on in Switzerland in February 2025. With referenda taking place up to four times a year, Swiss (semi-)direct democracy represents a unique case for studying how citizens use search engines to seek political information in order to make voting decisions, and how this shapes different forms of information exposure.

Specifically, we test the assumption that users may be less susceptible to selective exposure when formulating search terms themselves than when exposed to a curated list of queries (e.g., as in the case of autocomplete suggestions recommended by search engines). To this end, we asked respondents about the potential use of search queries in both open-ended and closed-ended questions. We also investigate how pre-existing knowledge about the initiative, beliefs about climate change, political attitudes, and cognitive factors (in particular, more analytical vs. more intuitive thinking styles) affect selective exposure via search engines.

Importantly, we do not find direct evidence of selective exposure—either in self-formulated or in recommended search queries. In particular, respondents who intend to vote for or against the initiative are not biased in their search query preferences in either direction. Yet, we find that the expected vote outcome (i.e., the expectation regarding whether the initiative will be rejected or accepted by the majority) has a limited effect on query selection. We also find that individual differences, including demographics, climate beliefs, and cognitive factors, affect the subtopic of the query.

This study makes several contributions to the field of political communication. First, we show how individual characteristics of voters (e.g., knowledge, political attitudes, and cognitive factors) may affect information-seeking behavior on search engines and discuss how it can influence democratic decision-making. Second, we demonstrate that selective exposure, contrary to the results of other studies, is not necessarily present in online search and discuss how the political and media environment may affect this. Finally, we compare these findings with another survey wave conducted prior to an earlier round of popular votes on retirement policies, thus demonstrating the differences in information-seeking behavior on different political issues.

References

Ekström, A. G., Madison, G., Olsson, E. J., & Tsapos, M. (2024). The search query filter bubble: Effect of user ideology on political leaning of search results through query selection. *Information, Communication & Society, 27*(5), 878–894. https://doi.org/10.1080/1369118X.2023.2230242

Ekström, A. G., Niehorster, D. C., & Olsson, E. J. (2022). Self-imposed filter bubbles: Selective attention and exposure in online search. *Computers in Human Behavior Reports, 7*, 100226. https://doi.org/10.1016/j.chbr.2022.100226

van Hoof, M., Meppelink, C. S., Moeller, J., & Trilling, D. (2024). Searching differently? How political attitudes impact search queries about political issues. *New Media & Society, 26*(7), 3728–3750. https://doi.org/10.1177/14614448221104405

Robertson, R. E., Green, J., Ruck, D. J., Ognyanova, K., Wilson, C., & Lazer, D. (2023). Users choose to engage with more partisan news than they are exposed to on Google Search. *Nature, 618*(7964), 342–348. https://doi.org/10.1038/s41586-023-06078-5

Vziatysheva, V., Makhortykh, M., Sydorova, M., & Jumle, V. (2024). Google, how should I vote? How users formulate search queries to find political information on search engines. *arXiv*. https://doi.org/10.48550/arXiv.2410.00778

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Sensemaking in the Age of Algorithms

Young Users' Algorithmic Interactions and Agency Negotiation on TikTok

Leonie Wunderlich \\ Leibniz-Institute for Media Research | Hans-Bredow-Institute \\ Hamburg \\ Germany
l.wunderlich@leibniz-hbi.de

Sascha Hölig \\ Leibniz-Institute for Media Research | Hans-Bredow-Institute \\ Hamburg \\ Germany
s.hoelig@leibniz-hbi.de

Philipp Kessling \\ Leibniz-Institute for Media Research | Hans-Bredow-Institute \\ Hamburg \\ Germany
p.kessling@leibniz-hbi.de

KEYWORDS
algorithmic literacy; social media affordances; young people; sensemaking; agency negotiation; algorithmic recommender systems

Abstract

The AI-driven algorithm that powers TikTok's content recommendation system and platform affordances shape user behavior while users' perceptions of specific affordances recursively influence algorithmic outputs. Thus, users are both subjects to and co-creators of algorithmic structures within a constant interaction process. Based on Giddens's (1984) structuration theory as well as Oeldorf-Hirsch and Neubaum's (2023) algorithmic literacy framework, our study takes a user-centric approach for investigating this complex user-algorithm interaction. In a first step, six group discussions (n=31) with German 16- to 24-year-olds explored collective cognitive and affective reflections on TikTok's algorithmic recommender system (RQ 1+2). In a second step, we conducted follow-up interviews with two participants of each group (n=12). Visualizations of their individual TikTok Data Download Packages were used as a stimulus to investigate how participants engage platform-specific affordances in their interaction with the ARS (RQ 3). By combining qualitative insights with user-provided behavioral data, we produce nuanced findings that also inform journalism practices and educational efforts aimed at improving algorithmic literacy and news competence to empower young people's agency in online environments.

Introduction

The AI-driven algorithm that powers TikTok's content recommendation system and platform-specific affordances significantly determine how young users encounter and interact with content on the platform. Regarding informational and political content, studies show that news is relatively sparse in users' feeds, partly because the content recommendation algorithm that generates TikTok's "ForYou"-feed responds slightly to active signals of news interest from simulated users (Hagar & Diakopoulos, 2023). However, young people increasingly consider TikTok as a significant source of news (Behre et al., 2025; Granow & Scolari, 2022; Feierabend et al., 2024) and for political information consumption (McClain, 2024). Moreover, young users perceive TikTok's recommender system as highly effective and convenient, leading to heavy reliance on their curated ForYou-Page (Kang & Lou, 2022; Narayanan, 2023). For example, qualitative studies reveal that teenagers and young adults spend most of their time on TikTok's FYP, because the algorithm is perceived as "very good" and perfectly tailored to their interests (NRW Media authorities, 2024). Schellewald (2023, p. 1579) also shows how young people "imagined TikTok as affording convenient access to relatable content that catered to their escapist desires and needs".

While young adults accept the role of algorithm-driven curation in guiding content consumption on TikTok, they also play an active role in the user-algorithm-relationship by engaging with the algorithm (Schober et al., 2022) and platform affordances (Schellewald, 2023) to align the algorithm more precisely to their needs (Kang & Lou, 2022, p. 1). On a cognitive level, however, young people often lack the ability to articulate mechanisms behind news personalization, and mere knowledge about algorithms does not necessarily empower them to intervene in algorithmic decision-making (Swart, 2021). Rather, users find themselves in a constant process of agency negotiation on TikTok, meaning to "be guided by the machine and conform to its directives or exert control over it by customizing settings" (Sundar, 2020, p. 82).

While most studies focus on the cognitive aspects of algorithmic media use (Cotter & Reisdorf, 2020; Dogruel, Masur & Joeckel, 2021; Gagrčin, Naab & Grub, 2024), little is known about the affective dimension (Oeldorf-Hirsch & Neubaum, 2023) and its connection to other stages in acquiring algorithm literacy (Gagrčin et al., 2024). Regarding the behavioral aspects of how users practically engage with algorithms (Kang & Lou, 2022; Schellewald, 2023), it remains unclear how platform specific affordances shape user-algorithm interaction. Against this background, our study takes a user-centric approach for investigating the complex interplay between users, algorithms, and media practices. We want to address the following research questions:

RQ1: In how far do young people reflect on the algorithmic recommender system (ARS) of TikTok and what do they know about it?

RQ2: What are young users' affective responses, perceived benefits, and concerns of it?

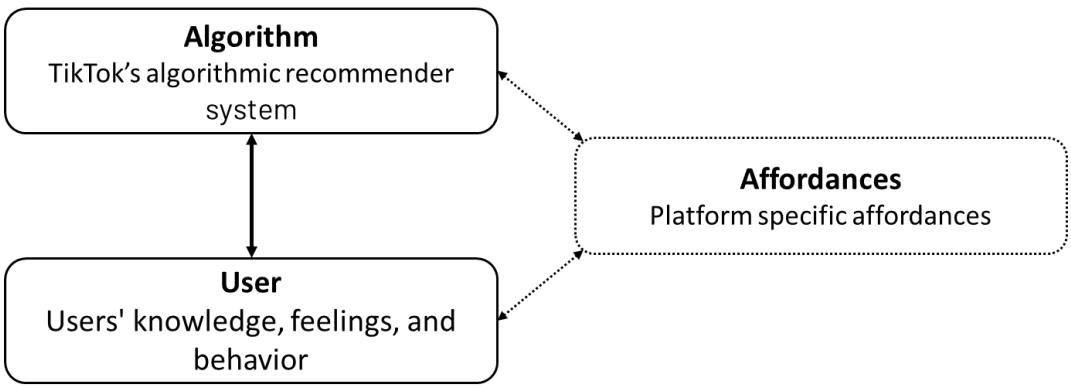
RQ3: In how far do young people use platform-specific affordances in their interaction with the ARS and how do they evaluate their agency within this process?

The goal is to gain more insights into the interplay and dynamics between affective, cognitive, and behavioral aspects influencing the user-algorithm interaction and agency negotiation processes with algorithmic recommender systems of social networking platforms.

Theoretical Foundation

Giddens's (1984) structuration theory offers a useful theoretical framework for understanding this dynamic because it allows for the integration of both individual level and macro level constructs. Structuration theory is led by the three major elements: *agents* (users), *structures* (algorithmic structure), and *duality*. The latter highlights that structures are both the result and condition of human actions. In the context of TikTok, an "AI-based algorithmic social networking site" (Kang & Lou, 2022, p. 4), the algorithmic recommender system that is shaped by platform-specific affordances forms the structure. On the individual level, user behavior is accompanied by cognitive and affective aspects (Oeldorf-Hirsch & Neubaum, 2025). User actions afforded by the algorithmic features of the platform lead to different types of experiences and user-algorithm interactions. For example, the Like button affords the signaling of 'liking' content. This technical feature thus enables the expression of emotion and is therefore an important component of young users' communication and participation practices on social media platforms (Olsson, 2016). In the user-algorithm relationship, this function can also be seen as an expression of engagement with the algorithm (Klug et al., 2021). For example, users "like" certain content to get more thematically similar content displayed on the FYP (Schober et al., 2022) and regularly "train" the algorithm to show desirable videos (Siles & Meléndez-Moran, 2021). How users deal with and use platform-specific affordances depends on their intention, perception, and knowledge as well as the social context of the usage situation (Bucher & Helmond, 2018). Applied to the context of our study, the aspect of duality manifests in the algorithmic structure shaping user behavior while user perceptions of affordances understood as "possibilities for actions" (Evans et al., 2017, p. 36) recursively influence algorithmic outputs (Kang & Lou, 2022; Obreja, 2024; Swart, 2021). Thus, users are both subjects to and co-creators of algorithmic structures within a constant interaction process (see Figure 1).

Figure 2: Framework of user-algorithm interaction

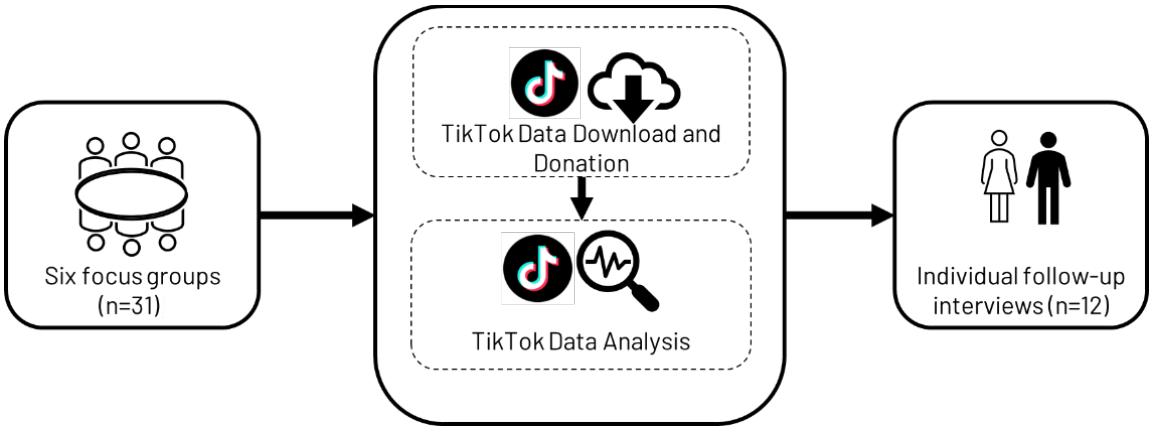


Method

To comprehensively examine the user-algorithm interaction dynamics and to answer the research questions, focus groups are combined with in-depth interviews in which we also look at individual TikTok data donated by the users (see Figure 2). Before recruitment, ethical questions and consent were discussed and obtained in close consultation with the ethics board to address the sensitivity involved in working with minors. For safeguarding ethical research practices and participants' privacy we followed van Driel et al. (2022) four-point approach.

Participants were recruited by Krämer market research institute according to predefined criteria. Two focus groups were held in person in three major cities Hamburg, Düsseldorf, and Erfurt, each with a mixed composition in terms of age, gender, and formal education. Data collection for the focus groups took place in March and April, the individual interviews were conducted in May and June, mostly online.

Figure 3: Mixed-methods design of the study



In a first step, in six group discussions with young participants (n= 31), collective knowledge, shared experiences, and general attitudes toward TikTok's algorithmic personalization as well as TikTok's role in news and political information consumption were explored (RQ 1+2). Here, participants had to work together on a group assignment "Your ForYou-Page as your favorite dish" on factors they assume to influence the recommended content on their FYP. First, they worked individually to write down the ingredients (influencing factors) needed to create their favorite dish (FYP). They then discussed this as a group, identified similarities and differences and named the main categories for the factors and categorized their individual pieces of paper accordingly. The aim was to observe the assumptions about influencing factors and the articulation of the participants' experiences and knowledge.

In the second step, we conducted follow-up interviews with two participants of each focus group (n=12). After the focus groups, they received instructions on how to download and donate their TikTok data. Analyzing the TikTok data, we looked at a period of three to four months, which included the federal election in Germany in February 2025. No personal data was processed, but interaction data such as likes, comments, shared content, search terms, followed and blocked accounts as well

as the number of videos watched. The visualized data on participants' TikTok interactions was used as a stimulus in the interviews. Hence, it served a) for validation and/or comparison with self-reported behavior, b) reflection and evaluation of certain interaction practices afforded by the platform c) as anchor points for describing and understanding various stages within the user-algorithm interaction by using concrete examples such as search terms entered.

Regarding data analysis, we want to apply different approaches for the focus group and interview data. For example, for the discussion we want a more open coding approach to define themes that evolve during the group task whereas for the interview data we could start with a predefined code list. For the latter, we could use the cognitive, affective, and behavioral dimensions of algorithmic literacy as classified by Gagrčin and colleagues (2024) for main and subcategories. Overall, we aim to systematically relate both data sets to each other within the theoretical framework.

Preliminary Results and Discussion

As the data collection was only completed in June, the analysis and presentation of the results of the study is still in progress. Initial findings from the focus groups indicate that all participants have a high level of superficial algorithmic awareness that is displayed in in the prompt answers on the question of what content is consumed on TikTok: "what the algorithm shows me", "that's determined by the algorithm" or "the algorithm knows what I want to see". Participants' knowledge of different criteria that are assumed to influence the algorithmic recommender system varied by group. There are age-related differences between younger and older groups regarding the number of distinct factors and the degree of differentiation of the top-level categories (between three and five main categories per group). Overall, focus groups participants' reflections of algorithms seem to be based more on personal experiences and interactions (behavioral) than on actual knowledge about how algorithms in general and content recommendation algorithms in particular work (cognitive).

This study integrates individual-level user behavior with macro-level algorithmic affordances to explore the bidirectional nature of user-algorithm interaction. By integrating qualitative insights with user-provided behavioral data, we will produce nuanced findings that also inform journalism practices and educational efforts aimed at improving algorithmic literacy and news competence to empower young people's agency within online environments.

References

Behre, J., Hölig, S., Stöwing, E., & Möller, J. (2025). *Reuters Institute Digital News Report 2025 – Ergebnisse für Deutschland* (Arbeitspapiere des Hans-Bredow-Instituts No. 77). Verlag Hans-Bredow-Institut. https://doi.org/10.21241/ssoar.102887

Bucher, T., & Helmond, A. (2018). The affordances of social media platforms. In J. Burgess, A. Marwick, & T. Poell (Eds.), *The SAGE handbook of social media* (pp. 233–235). SAGE.

Dogruel, L., Masur, P., & Joeckel, S. (2021). Development and validation of an algorithm literacy scale for internet users. *Communication Methods and Measures, 16*(2), 115–133. https://doi.org/10.1080/19312458.2021.1968361

Evans, S. K., Pearce, K. E., Vitak, J., & Treem, J. W. (2017). Explicating affordances: A conceptual framework for understanding affordances in communication research. *Journal of Computer-Mediated Communication, 22*(1), 35–52. https://doi.org/10.1111/jcc4.12180

Feierabend, S., Rathgeb, T., Gerigk, Y., & Glöckler, S. (2024). *JIM 2024. Jugend, Information, Medien. Basisuntersuchung zum Medienumgang 12- bis 19-Jähriger in Deutschland*. Medienpädagogischer Forschungsverbund Südwest. https://mpfs.de/app/uploads/2024/11/JIM_2024_PDF_barrierearm.pdf

Gagrčin, E., Naab, T. K., & Grub, M. F. (2024). Algorithmic media use and algorithm literacy: An integrative literature review. *New Media & Society*. https://doi.org/10.1177/14614448241291137

Giddens, A. (1984). The constitution of society: Outline of the theory of structuration. University of California Press.

Granow, V. C., & Scolari, J. (2022). Ergebnisse einer Mixed-Methods-Grundlagenstudie im Auftrag des SWR. TikTok – Nutzung und Potenziale der Kurzvideo-Plattform. *Media Perspektiven*, (4), 166–176.

Hagar, N., & Diakopoulos, N. (2023). Algorithmic indifference: The dearth of news recommendations on TikTok. *New Media & Society*. https://doi.org/10.1177/14614448231192964

Kang, H., & Lou, C. (2022). AI agency vs. human agency: Understanding human–AI interactions on TikTok and their implications for user engagement. *Journal of Computer-Mediated Communication, 27*(5), 1–13. https://doi.org/10.1093/jcmc/zmac014

Landesanstalt für Medien NRW. (2024). *Ergebnisbericht: Tagebuch-Studie. Bedeutung von TikTok und Instagram als politische Informationsmedien für junge Menschen*. https://www.medienanstalt-nrw.de/fileadmin/user_upload/Forschung/Online-Tagebuchstudie_Ergebnisbericht.pdf

Obreja, D. M. (2024). When stories turn institutional: How TikTok users legitimate the algorithmic sensemaking. *Social Media + Society, 10*(1). https://doi.org/10.1177/20563051231224114

Oeldorf-Hirsch, A., & Neubaum, G. (2025). What do we know about algorithmic literacy? The status quo and a research agenda for a growing field. *New Media & Society, 27*(2), 681–701. https://doi.org/10.1177/14614448231182662

Olsson, T. (2016). Social media and new forms for civic participation. *New Media & Society, 18*(10), 2242–2248. https://doi.org/10.1177/1461444816656338

Schellewald, A. (2023). Understanding the popularity and affordances of TikTok through user experiences. *Media, Culture & Society, 45*(8), 1568–1582. https://doi.org/10.1177/01634437221144562

Schober, M., Lauber, A., Bruch, L., Herrmann, S., & Brüggen, N. (2022). „Was ich like, kommt zu mir“: Kompetenzen von Jugendlichen im Umgang mit algorithmischen Empfehlungssystemen. Qualitative Studie im Rahmen von „Digitales Deutschland“. *JFF – Institut für Medienpädagogik in Forschung und Praxis*. München: kopaed.

Siles, I., & Meléndez-Moran, A. (2021, May). “The most aggressive of algorithms”: User awareness of and attachment to TikTok’s content personalization. Paper presented at the *71st Annual Conference of the International Communication Association (ICA)*, Denver, USA.

Sundar, S. S. (2020). Rise of machine agency: A framework for studying the psychology of human–AI interaction (HAII). *Journal of Computer-Mediated Communication, 25*(1), 74–88. https://doi.org/10.1093/jcmc/zmz026

Swart, J. (2021). Experiencing algorithms: How young people understand, feel about, and engage with algorithmic news selection on social media. *Social Media + Society, 7*(2), 1–11. https://doi.org/10.1177/20563051211008828

van Driel, I., Giachanou, A., Pouwels, J. L., Boeschoten, L., Beyens, I., & Valkenburg, P. M. (2022). Promises and pitfalls of social media data donations. *Communication Methods and Measures, 16*(4), 266–282. https://doi.org/10.1080/19312458.2022.2109608

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

On the Relationship of Digital Well-Being and Digital Literacy

Laura Sūna \\ University of Siegen \\ Siegen \\ Germany
laura.suna@uni-siegen.de

Dagmar Hoffmann \\ University of Siegen \\ Siegen \\ Germany
dagmar.hoffmann@uni-siegen.de

KEYWORDS
digital literacy; well-being; survey

Abstract

What do people in mediatized societies need in order to lead a "good life"? This question has both socio-political and individual relevance – especially in times of multiple crises. In the context of the establishment of digital media technologies in everyday life and many areas of society, current research literature defines the "good life" as the well-being of media users. It examines how digital media can improve or impair lives. However, empirical studies are quite ambivalent: on the one hand, a negative influence of the digital media technologies on the well-being of users has been proven, but on the other hand, positive effects have also been identified (Kannengießer 2022). It is evident that media users weigh up the advantages and disadvantages of digital media in their everyday lives depending on the situation and routinely adapt their behavior to these considerations (Sūna & Hoffmann 2024).

Studies on digital well-being address the need for development of the ability of neutralizing the side effects of digital communication concerning problems in the management of connected digital devices in people's daily life. Beyond a pathological dimension (e.g. "internet addiction", "digital exhaustion/fatigue"), more and more studies show that the majority of internet users suffers from problems coping with communication overload both at work and in their private lives (Nitsch & Kinnebrock 2023; Gui et al. 2017). This has been summarized as digital overconsumption and touches questions about how and why people feel they are using digital media more than they would like to. Another issue is that of multi-tasking, which identifies the condition of a continuous switching between different focuses of attention on digital media. These two problems occur

simultaneously and are inextricably linked. Algorithmic recommendations and other automated systems reinforce the pull effect of social media in particular.

Gui et al. (2017) define "digital well-being" as a state where subjective well-being is maintained in an environment characterized by digital communication overabundance. Resilient individuals can channel digital media usage towards a sense of comfort, safety, satisfaction and fulfilment. Specific individuals' skills and also the socio-cultural context they live in favour this situation. These media users are able to cope with the flipside effects of digital media while using them to obtain a wide range of benefits. It can be achieved by strengthening digital literacy of citizens. Following digital skills can enable a digital well-being in a mediatized society: instrumental skills (the skills necessary to use digital media), cognitive and critical-reflexive skills (knowledge about digital media and how to evaluate them), creative skills (concerning the self-determined (re)design of digital media and systems), and affective and social skills (being able to react emotionally and socially appropriately to media content and systems) (Digitales Deutschland 2021). There is consensus that specific competence requirements are placed on subjects in order to participate in a democratic society characterized by deep mediatization (Hepp 2020). The phenomenon of digital well-being illustrates that individual skills and culturally anchored values and norms always belong together.

Based on representative survey data from three points in time (2021, 2023, 2025), the paper describes how media users in Germany evaluate their own digital skills and how they evaluate the significance of different skills in the context of societal future and digital well-being. Overall, the data shows a slight discrepancy between the self-assessment of one's own media-related skills and the assessment of the importance of these skills for society. Of course, the majority of respondents want to use digital media in a way that is good for them. At the same time, however, they state that they are not always able to set limits on their own media use. Most of the respondents see a particular need for support in those aspects of digital literacy that refer to social norms and values. Aspects such as trust in online sources, risk assessment on the internet and the assessment of credibility online are considered important by the respondents and they would like to see further education opportunities in these areas. Social responsibility skills in online spaces are also relevant for the respondents (Cousseran et al. 2023). Consistent with the research of Livingstone et al. (2021/2023), our findings confirm that higher digital literacy does not necessarily minimize perceptions of safety and online risks and increase well-being and guarantees a "good life".

References

Cousseran, L., Lauber, A., Herrmann, S., & Brüggen, N. (2023). *Kompass: Künstliche Intelligenz und Kompetenz 2023. Einstellungen, Handeln und Kompetenzentwicklung im Kontext von KI*. kopaed. https://doi.org/10.5281/ZENODO.10058588

Digitales Deutschland. (2021). *Digitales Deutschland: Ein Rahmenkonzept*. https://digid.jff.de/wp-content/uploads/2021/06/Rahmenkonzept_DigitalesDeutschland_Vollversion.pdf

Gui, M., Fasoli, M., & Carradore, R. (2017). “Digital well-being”: Developing a new theoretical tool for media literacy research. *Italian Journal of Sociology of Education, 9*(1), 155–173. https://doi.org/10.14658/pupj-ijse-2017-1-8

Hepp, A. (2020). *Deep mediatization*. Routledge.

Kannengießer, S. (2022). *Digitale Medien und Nachhaltigkeit: Medienpraktiken für ein gutes Leben*. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-36167-9

Livingstone, S., Mascheroni, G., & Stoilova, M. (2021). The outcomes of gaining digital skills for young people’s lives and wellbeing: A systematic evidence review. *New Media & Society, 25*(5), 1176–1202. https://doi.org/10.1177/14614448211043189

Nitsch, C., & Kinnebrock, S. (2023). “I urgently need your advice”—Digital stress experiences and social support in online forums. *International Journal of Communication, 17*, 4037–4056. https://ijoc.org/index.php/ijoc/article/view/19523/4217

Sūna, L., & Hoffmann, D. (2024). From AI imaginaries to AI literacy: Artificial intelligence technologies in the everyday lives of migrants in Germany. *MedieKultur: Journal of Media and Communication Research, 40*(76), 53–76. https://doi.org/10.7146/mk.v40i76.137144

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Autonomy and Relationality in the Digital Age

Eloïse Soulier \\ University of Hamburg \\ Hamburg \\ Germany
eloise.soulier@uni-hamburg.de

KEYWORDS
relational autonomy; conceptual ethics

Abstract

Guidelines and legal frameworks on artificial intelligence and digital technologies often consider human autonomy as one of the main values to protect (Jobin, Ienca, and Vayena 2019). Abundant academic scholarship is also concerned with autonomy and digital technologies (Susser, Roessler, and Nissenbaum 2019; Yeung 2019; Rubel, Castro, and Pham 2021). The concern for autonomy is neither new nor limited to digital technologies, in fact autonomy arguably underlies political liberalism and as such, has been the object of abundant literature. What is specific to the case of digital technologies is the wide-ranging scope, but also untransparent and highly adaptive nature of these technologies, which make this worry particularly salient (Susser, Roessler, and Nissenbaum 2019). This concern for human autonomy is related to the protection of individual freedom and dignity, but also to political concerns for electoral processes, fueled by numerous cases – or suspicions – of election manipulation[1].

Western law in general, including the regulation of digital technologies, arguably relies on a Kantian understanding of autonomy as self-legislation by rational atomistic individuals (Nedelsky 2011; Zarsky 2019). In this understanding, respecting a user's autonomy means providing them with all potentially relevant information and then not interfering with their decision-making process. In the context of digital technologies, this is exemplified by regulatory measures such as informed consent approaches to data protection. I argue that the widely discussed limitations of informed consent approaches illustrate the failure of a Kantian conception of autonomy to live up to its own ideal of rational independent decision-making. It can even be argued to rather serve a neoliberal agenda

1 See e.g. https://www.nytimes.com/2017/12/11/opinion/fake-news-russia-kenya.html, https://www.theguardian.com/uk-news/2018/apr/04/cambridge-analytica-used-violent-video-to-try-to-influence-nigerian-election

that overburdens individuals, while releasing corporate actors from responsibility. In fact, scholarly critiques of the Kantian concept of autonomy, coming from different strands of scholarship, have argued that this concept is epistemically and normatively flawed. Integrating these critiques is crucial to develop a better understanding – and therefore being better able to protect what it is that we seem to value so much about autonomy.

These critiques of the Kantian concept of autonomy pertain both to its epistemic value and to the normative ideal it serves. Perhaps most famously, psychology and the cognitive sciences have struck a blow at the idea of rational decision making, highlighting the different ways in which we deviate from this ideal in our practical decision-making (Kahneman and Tversky 1977). On the other hand, critical theories, importantly feminist theory, have challenged this conception as obscuring our relationality and interdependence, and in so doing conveying a masculinist, individualist ideal of personhood (Mackenzie and Stoljar 2000). Critiques stemming from the philosophy of technology, but also from disability studies, have underlined how we are dependent on, if not even determined by, the technological infrastructure (Verbeek 2011, Ells 2001). Scholars who challenge this conception of autonomy however usually recognize the need for *a* concept of autonomy, crucially due to the necessity for any emancipatory project to have a concept of autonomy (Nedelsky 2011). The question then is rather: which concept of autonomy? Within the framework of pragmatist conceptual ethics that I draw from (Thomasson 2020), this question amounts to examining the function fulfilled by the concept of autonomy. What purposes do we want this concept to serve, what kind of ideal representation of selves should it be conveying? Indeed, underlying the argument of this article is the assumption that our conceptual choices both reflect and perpetuate normative commitments, and therefore conceptual discussions are essential to the debate on the kind of digital experiences we want to have.

In this framework, drawing from feminist scholarship is very fruitful, since feminist scholars have historically been more aware of and explicit about engaging in theorizing that serves a purpose, namely the purpose overcoming oppression. For this article, I draw on relational accounts of autonomy (see Mackenzie and Stoljar 2000 for an overview). I argue that adopting a conception of autonomy that makes room for relationality is not only epistemically and practically more fruitful, but also normatively more desirable. Nevertheless, drawing on work by John Christman (2004) and Serene Khader (2020), I claim that emancipation, which is an essential purpose of autonomy for critical theories, requires a conception of autonomy not to be constitutively relational. Aiming to reconcile the partly conflicting constraints set by these two essential functions of the concept of autonomy, I propose to operate with a conception of autonomy as "the ability to structure our dependences". This reconceptualization arguably does not provide an answer to all the debates on the concept of autonomy, nor is it my ambition. My aim is to shift the focus, when thinking about user autonomy and digital technologies, from the perfectly informed atomistic decision-maker to appropriate structures of delegation.

Finally, I show that these rather theoretical reflections on the concept autonomy can have very practical consequences for the design and regulation of digital technologies, of which I discuss two examples. The first is the PIMS (personal information management systems) mechanisms for data protection, that would theoretically enable the kind of organized and voluntary delegation that

operationalize the proposed conception of autonomy. The second example is the design of recommender systems of all kinds in a way that supports different forms of chosen dependence, for instance allowing users to orient on what a trusted source, friends, or a set of principles would recommend. Through these examples, I hope to show that conceptual discussions – and negotiations – on the values that we want our concepts to serve, can have very concrete consequences on our digital experiences and wellbeing. Our meanings are in constant transformation, they are always being politically negotiated. But digital technologies embed concepts and the worldviews they carry with them in a way that both rigidifies and multiplies them. Hence the essential need to reflect on what these concepts should mean. To end with Haslanger's apt words: "We should (at least try to) take control over meanings, for if we don't, others will" (Haslanger 2020).

References

Christman, J. (2004). Relational autonomy, liberal individualism, and the social constitution of selves. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 117*(1/2), 143–164.

Ells, C. (2001). Lessons about autonomy from the experience of disability. *Social Theory and Practice, 27*(4), 599–615.

Haslanger, S. (2020). Going on, not in the same way. In A. Burgess, H. Cappelen, & D. Plunkett (Eds.), *Conceptual engineering and conceptual ethics* (pp. 230–260). Oxford University Press.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399.

Kahneman, D., & Tversky, A. (1977). Intuitive prediction: Biases and corrective procedures (Tech. Rep.). Decisions and Designs Inc.

Khader, S. J. (2020). The feminist case against relational autonomy. *Journal of Moral Philosophy, 17*(5), 499–526.

Mackenzie, C., & Stoljar, N. (Eds.). (2000). Relational autonomy: Feminist perspectives on autonomy, agency, and the social self. Oxford University Press.

Nedelsky, J. (2011). Law's relations: A relational theory of self, autonomy, and law. Oxford University Press.

Rubel, A., Castro, C., & Pham, A. (2021). *Algorithms and autonomy: The ethics of automated decision systems*. Cambridge University Press.

Susser, D., Roessler, B., & Nissenbaum, H. (2019). Technology, autonomy, and manipulation. *Internet Policy Review, 8*(2). https://doi.org/10.14763/2019.2.1410

Thomasson, A. (2020). A pragmatic method for normative conceptual work. In A. Burgess, H. Cappelen, & D. Plunkett (Eds.), *Conceptual engineering and conceptual ethics* (pp. 435–458). Oxford University Press.

Verbeek, P. P. (2011). Subject to technology: On autonomic computing and human autonomy. In M. Hildebrandt & A. Rouvroy (Eds.), *Law, human agency and autonomic computing* (pp. 27–45). Routledge.

Yeung, K. (2019). 'Hypernudge': Big data as a mode of regulation by design. In M. Hildebrandt (Ed.), *The social power of algorithms* (pp. 118–136). Routledge.

Zarsky, T. (2019). Privacy and manipulation in the digital age. *Theoretical Inquiries in Law, 20*(1), 157–188.

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Panel 3

Session 7: Civic Engagement and Inclusion

Session 8: Participation and Co-Production

Session 9: Trust and Security

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Making a Good Society Through a Good Life

The Maker Movement, Individual Eudaimonia, and Democratic Participation

Martin Rolf Herbers \\ Zeppelin Universität \\ Friedrichshafen \\ Germany
martin.herbers@zu.de

KEYWORDS
maker movement; public sphere; political participation; media practice; eudaimonia

The Maker Movement as an Interconnector of Individual and Collective Practices

This paper analyzes the eudaimonic and democratic aspects of the Maker Movement across three dimensions: individual actors, institutional politics, and their interconnections through "publicness" (Herbers, 2024). It argues that individual Makers, driven by eudaimonic motivation to critically and creatively engage with technology, also contribute to democratic practice through "material participation" (Marres, 2015). Thus, personal engagement with technology extends beyond the individual, linking to broader democratic society.

To support this argument, the paper offers a theoretical reflection on the Maker Movement's individual and collective aims. It presents original empirical findings on "making" practices, supplemented by secondary research. From these analyses, the paper proposes integrating individual eudaimonic technological engagement into theories of democratic participation, positioning it at the center of digital societies.

The Goals and Aims of the Maker Movement

The Maker Movement is a global initiative fostering a critical understanding of technology (Dougherty, 2012). Members dismantle and redesign technologies to enhance literacy and personal agency (Dougherty, 2016).

Beyond technological engagement, the movement promotes consumption-critical media practices as informal political participation (Kannengießer, 2016). Through "material participation" (Marres, 2015), Makers influence both individual agency and broader social and political landscapes. As "making is making in public" (Anderson, 2012), they engage in digital and physical spaces to enhance technological awareness and accessibility (Gauntlett, 2018).

Collectively, they contribute to policymaking on STEAM education, sustainability, and entrepreneurship, evidenced by their engagement in institutional politics, including a 2008 White House gathering (Miller, 2014). They are also shaped by policy frameworks, as seen in EU initiatives supporting the Maker Movement (Rosa et al. 2017).

Eudaimonia and Democratic Participation

The Maker Movement is driven by eudaimonic motivation at the individual level. As Oliver and Bartsch (2011) argue, eudaimonia arises from engaging in practices that go beyond immediate gratification, fostering long-term personal growth through reflection (Wirth, Hofer, & Schramm, 2012). Key drivers of eudaimonic practice include personal autonomy and social connection (Rieger et al. 2014). Makers, as a "pioneer community," engage with technology not only out of passion but also to explore its societal implications (Hepp, 2016).

Eudaimonic practices extend beyond individuals, influencing democratic societies by fostering technological agency, which in turn enhances democratic participation. Making, as a practice, establishes a form of "publicness" (Herbers, 2024), linking personal engagement with collective social structures.

Critical Perspectives

While the Maker Movement promises to democratize technological engagement, critical perspectives reveal structural limitations. Social inclusion does not automatically follow from infrastructural openness. Instead, a "maker elite" emerges, a predominantly male, technologically proficient, and socially homogeneous group, contradicting the inclusive ideals of the movement (Ferretti & van Lente, 2022). These findings urge scholars to reconsider the romanticization of making as inherently democratic and instead interrogate its normative blind spots: Who gets to make? Who remains excluded? And how do local cultures shape or constrain the publicness of participation?

Implications for Research

This paper proposes a conceptual shift: From "public discourse" to "material publicness" as a framework to understand democratic participation in digital societies. Future research should empirically investigate how technological agency manifests across different social groups, and under which conditions eudaimonic practices translate into civic engagement. Comparative studies across urban and rural Makerspaces, as well as analyses of gendered and class-based access patterns, can further clarify the democratic potentials and pitfalls of these spaces. Additionally, research should examine how policy frameworks (e.g., EU funding instruments) shape the institutionalization of making—and whether this supports or hinders the development of truly inclusive technological publics.

References

Anderson, C. (2012). *Makers: The new industrial revolution*. Crown Business.

Chidgey, R., & Marotta, A. (2021). Feminist hackerspaces as repair labs of care. *Feminist Media Studies, 21*(3), 370–387.

Dougherty, D. (2012). The maker movement. *Innovations: Technology, Governance, Globalization, 7*(3), 11–14.

Dougherty, D. (2016). Free to make: How the maker movement is changing our schools, our jobs, and our minds. North Atlantic Books.

Ferretti, F., & Van Lente, H. (2022). The promise of the maker movement: Policy expectations versus community criticisms. *Science and Public Policy, 49*(1), 18–27.

Fraser, N. (1990). Rethinking the public sphere: A contribution to the critique of actually existing democracy. *Social Text, (25/26)*, 56–80.

Gauntlett, D. (2018). Making is connecting: The social power of creativity, from craft and knitting to digital everything (2nd ed.). Polity Press.

Hepp, A. (2016). Pioneer communities: Collective actors in deep mediatisation. *Media, Culture & Society, 38*(6), 918–933.

Kannengießer, S. (2016). Practices of media and material participation: A theoretical framework. *Media and Communication, 4*(2), 36–47.

Marres, N. (2015). Why political ontology must be experimentalized: On eco-show homes as devices of participation. *Social Studies of Science, 45*(3), 311–334.

Rieger, D., Reinecke, L., Frischlich, L., & Bente, G. (2014). Media entertainment and eudaimonic enjoyment: Understanding the role of hedonic and eudaimonic needs in entertainment consumption. *Journal of Communication, 64*(3), 456–478.

Rosa, A., Ferretti, F., Perini, A., Sillitti, A., & Succi, G. (2017). *Enabling makerspaces: Recommendations for fostering innovation and education*. JRC Science for Policy Report. Publications Office of the European Union.

Wirth, W., Hofer, M., & Schramm, H. (2012). Beyond pleasure: Exploring the eudaimonic entertainment experience. *Human Communication Research, 38*(4), 406–428.

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Create Against Hate

Historical Learning and Digital Media Literacy for Playful Resilience in Algorithmic Environments

Tobias Ebbrecht-Hartmann \\ The Hebrew University of Jerusalem \\ Jerusalem \\ Israel
tobias.ebbrecht-hartmann@mail.huji.ac.il

KEYWORDS
media literacy; holocaust; antisemitism; digital education; TikTok; algorithm

Abstract

AI-generated videos about concentration camps, virtual avatars mocking Anne Frank, and algorithmically amplified misinformation about historical facts tremendously challenge efforts to educate and learn about difficult pasts on social media platforms like TikTok. The dangers posed by deepfakes, online propaganda, antisemitism, and algorithm-driven hatred on social media constitute a double-edged sword for participatory learning about history. This paper examines to what extent multimodal formats such as short videos can constitute "affinity spaces" (Gee 2017) in or beyond commercial, algorithm driven social media platforms. Focusing on educational accounts and videos this paper examines the complex role digital media play in shaping knowledge, discourse, and community engagement with regards to the specifically contested fields of learning about antisemitism and the Holocaust.

Introduction

AI-generated images and algorithmically distributed content about past events has significantly shaped the perception of history on platforms like TikTok and Instagram. Recently, AI-generated short videos presuming individual experiences of iconic historical events – ranging from Cleopatra in ancient Egypt to Anne Frank hiding from Nazi perpetrators – have flooded the For-You pages of millions of TikTok users. Most of these videos present history condensed in less than a minute long video collages of well-known and often stereotypically visualized past events. Instead of historical accuracy, they highlight the Point-of-View (POV) aesthetic, popular on short video platforms, and the astonishing potential of affordable and prompt-based tools that allow imagining and

generating any kind of (historical) reality. Some creators, however, also exploit the manipulative potential of AI-based image generation. A recent study of the Anne Frank Bildungsstätte in Frankfurt identified accounts that use AI-generated avatars to distort history and distribute antisemitic tropes and conspiracy theories (Schnabel & Berendsen 2025). As distribution of content on platforms like TikTok is based on tailored recommendation algorithms monitoring user behavior, such videos communicating disinformation specifically challenge learning about complex and difficult pasts like the history of the Holocaust.

To counteract such harmful aspects of digital media communication, it is crucial to promote access to content, formats, and practices that build resilience against hatred and exclusion while encouraging interactive and constructive social engagement. Digital education plays a central role in this effort, particularly in enhancing digital media literacy. This paper examines how digital formats can address complex societal challenges in ways that empower users and build resilience. Within this context, I focus on TikTok—an especially controversial and contested platform that not only exemplifies the ambivalent nature of social media as a learning space but also demonstrates the complexities of user engagement in an algorithm-driven environment. By uniquely combining user-generated content with algorithmic curation based on user tracking and recommendation systems, TikTok serves as a key site for exploring both the opportunities and risks of digital education (Siles et al. 2024).

Disinformation and Education on TikTok

Research indicates that social media platforms have become a suitable place for disinformation and propaganda. Lutz and Hoffmann (2017, 885) described this as “negative active participation”. The engaging environment and participatory practices of social media platforms like TikTok also allow users to “actively choose to engage for a purpose widely considered harmful or undesirable.” Antisemitic and far right actors have learned to exploit the algorithmic features and formats of social media platforms to their advantage, enabling them to distribute content to users who may not have previously engaged with far-right ideologies or conspiracy theories. One key strategy involves leveraging the compressed, memetic structure of short videos. Shifman (2014, 347) emphasizes that memes not only reflect but also actively shape “the ways of thinking, behaving, and acting of social groups.” In the age of online communication, memes became a regular part of political discourse on social media platforms. Far right actors use them to attract followers, build online communities and in particular for constructing unifying enemy stereotypes and identify targets for their intimidating propagandistic attacks (Trillo & Shifman 2021). This includes particular far-right and also antisemitic codes, which can be described as a “memification” of far-right and antisemitic disinformation (Ebbrecht-Hartmann 2024). As a result, democratic and historical-political education is being forced into a reactive and increasingly polarized mode of engagement. Instead of promoting shared learning and collaborative knowledge generation, pluralistic education is being stifled, along with the ability to explore controversial and diverse perspectives.

The counter-model to these forms of exclusive community building on platforms could be the development of affinity spaces around accounts and their followers or alternative platform environments, which not only react to forms of negative participation, but also provide a virtual space of empowerment through access to contextualizing information, historical and political education, and the promotion of critical media literacy. According to Gee, affinity spaces are "becoming prime spaces where people engage in 21st Century teaching, learning, doing and being." (Gee 2017, 28) They serve as physical as well as virtual social and cultural hubs where individuals connect, share ideas, and learn together based on mutual interests and passions (Gee 2018).

Based on these considerations, I examine how, in the field of historical education about the Holocaust and National Socialism, affinity spaces are established that reach beyond the passive sharing and absorption of information and encourage active engagement and promote users' skills, empowering them to engage critically and independently with history and complex mediated presents.

Digital Education as Playful Resilience

In the following, I explore how to build on and refine explanatory and learning approaches on TikTok to foster playful resilience against disinformation and manipulation within affinity spaces dedicated to antisemitism awareness and Holocaust history (Divon & Ebbrecht-Hartmann 2022).

With over 200,000 followers and nearly 111 million views, Susanne Siegert's German-language TikTok account *@keine_erinnerungskultur* is one of the most popular TikTok accounts dedicated to the Holocaust and National Socialism worldwide. A central element of Siegert's approach is her enduring presence in her videos. She frequently uses the POV perspective and directly addresses viewers, creating a personal and engaging experience This direct connection makes her a familiar and trusted figure for her followers—many users recognize her within the first few seconds of a video, knowing they can expect something engaging and insightful (Lormis 2023). Most of Siegert's videos are filmed in a private environment, often at her desk or kitchen table. In addition to sharing personalized stories about Nazi crimes and the biographies of victims, she also addresses right-wing symbols and codes on TikTok, right-wing extremism, and antisemitism.

A unique aspect of her videos is the encouragement for users to become active researchers themselves, exploring their own connections to the history of Nazi persecution. A prototypical video from January 2025 with 187.700 views, features Siegert sitting at her kitchen table, casually eating breakfast. In this intimate, everyday setting, she invites her audience to actively contribute to preserving the history of the Nazi past by collaborating in a collective effort to transcribe digitized evidence of Nazi persecution. She ties this effort to a challenge familiar to younger, social media-savvy generations: the increasing amount of passive screen time spent in front of digital devices. Acknowledging this, she reassures her followers that if they struggle with the same issue, "this video is for you." This direct address strengthens her connection with the audience. Rather than simply suggesting participation in a crowdsourcing project to transcribe historical records related to Nazi persecution,

Siegert demonstrates exactly how to do it, integrating screen recordings and walkthroughs into her video. This blends the familiar POV aesthetic with the instructional style of a video manual.

By engaging with traces of the past and analyzing historical sources, users enhance both their understanding of history and their digital media literacy. Through videos that demonstrate historical source criticism and expose contemporary forms of hate speech and antisemitism, Siegert empowers her audience to critically assess information and recognize patterns of misinformation.

Conclusion

As an affinity space for engaging with the history and memory of the Holocaust and Nazi persecution, this and other accounts, for instance by Holocaust memorials and museums (Ebbrecht-Hartmann & Divon 2024), offer an alternative approach to Holocaust education and commemoration—one that goes beyond declarations and rituals such as the widely used phrase "Never Again." Within the realm of digital education on social media, these creators and accounts respond to the platform's affordances while simultaneously injecting accurate and trustworthy content into an environment often shaped by misinformation and mistrust.

However, for this content to have a lasting impact, it must also be made accessible beyond the algorithm-driven structures of social media and further contextualized for digital education. One approach to create alternative ways of access to engage with short educational videos about the Holocaust and National Socialism is our recent project, *SHOAH STORIES: Learning about the Holocaust through Short Videos*, which aims to create a learning platform where teenagers can independently explore video content and accounts (shoahstories.video). In addition, the project offers educational approaches that provide educators with relatable, and engaging tools to integrate the short videos into their teaching, transforming the virtual affinity space into actual physical spaces of encounter and collaborative learning.

This shift is crucial in strengthening resilience against disinformation and the spread of polarizing narratives. It also addresses an increasing imbalance observed on commercial platforms like TikTok, where a small number of creators produce a disproportionate share of content (Bestvater 2024). To counteract this, it is essential to provide access to educational content beyond algorithm-driven distribution and to create creative spaces where individuals can develop the skills to explore and learn about complex historical topics in an engaging, factual, and responsible manner. This holistic approach to digital media enhances users' ability to think critically and navigate the evolving digital landscapes of information and disinformation.

Acknowledgement

Research for this paper was conducted as part of the project *SHOAH STORIES: Learning about the Holocaust with Short Videos* of the Anne Frank Zentrum Berlin funded by the Alfred Landecker Foundation.

References

Bestvater, S. (2024, February 22). How U.S. adults use TikTok: Around half of adult TikTok users in the U.S. have never posted a video themselves. And a minority of users produce the vast majority of content. Pew Research Center. https://www.pewresearch.org/internet/2024/02/22/how-u-s-adults-use-tiktok/

Divon, T., & Ebbrecht-Hartmann, T. (2022). Youthful platform commemoration: TikTok as a frontier for Holocaust education and memory. *Jewish Film & New Media, 10*(2), 231–249.

Ebbrecht-Hartmann, T. (2024). Memefizierter Antisemitismus: Protest und antisemitische Projektion auf TikTok, Instagram & Co im Schatten des 7. Oktobers. *CARS Working Papers, 21*. https://doi.org/10.17883/5058

Ebbrecht-Hartmann, T., & Divon, T. (2024). *Shoah commemoration and education on TikTok* [Report]. American Jewish Committee Berlin.

Gee, J. P. (2017). Affinity spaces and 21st century learning. *Educational Technology, 57*(2), 27–31.

Gee, J. P. (2018). Affinity spaces: How young people live and learn online and out of school. *Phi Delta Kappan, 99*(6), 8–13. https://doi.org/10.1177/0031721718762416

Lormis, J. (2023). #DoingMemoryOnTikTok: Gedenkstätten auf TikTok. Ist-Analyse und Perspektiven der Videoplattform TikTok für die Bildung und Vermittlung in Gedenkstätten [Master's thesis, Hochschule für Technik, Wirtschaft und Kultur Leipzig]. https://htwk-leipzig.qucosa.de/api/qucosa%3A87299/attachment/ATT-0/

Lutz, C., & Hoffmann, C. P. (2017). The dark side of online participation: Exploring non-, passive and negative participation. *Information, Communication & Society, 20*(6), 876–897. https://doi.org/10.1080/1369118X.2017.1293129

Schnabel, D., & Berendsen, E. (2025). Der Holocaust als Meme: Wie in digitalen Räumen Geschichte umgedeutet wird. Bildungsstätte Anne Frank.

Shifman, L. (2014). The cultural logic of photo-based meme genres. *Journal of Visual Culture, 13*(3), 340–358. https://doi.org/10.1177/1470412914546577

Siles, I., Valerio-Alfaro, L., & Meléndez-Morán, A. (2024). Learning to like TikTok... and not: Algorithm awareness as process. *New Media & Society, 26*(10), 5702–5718. https://doi.org/10.1177/14614448231176308

Trillò, T., & Shifman, L. (2021). Memetic commemorations: Remixing far-right values in digital spheres. *Information, Communication & Society, 24*(16), 2482–2501. https://doi.org/10.1080/1369118X.20201751866

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Combining Prebunking and Debunking Through Peer Production

The Case of CORRECTIV.Faktenforum

Caroline Lindekamp \\ CORRECTIV \\ Essen \\ Germany
caroline.lindekamp@correctiv.org

Anna Süß \\ CORRECTIV \\ Berlin \\ Germany
anna.suess@correctiv.org

KEYWORDS
disinformation; misinformation; participation; citizen journalism; fact-checking; media and information literacy (MILT);

Introduction

In an era where misinformation distorts public debate and manipulates opinions, fact-checking plays a crucial role in maintaining balanced discourse and providing orientation (Martel & Rand, 2024) in what Claire Wardle (2017) describes as the "Information Disorder." While debunking has been shown to be effective in addressing misinformation, the complexity and multi-causal nature of the problem requires an encompassing approach (Chan et al., 2017; Cook et al., 2017; Tay et al., 2022). Prebunking offers an additional preventive approach to preemptively strengthen societal resilience against manipulative narratives (Ecker et al., 2022). Media and information literacy training (MILT) has proven effective as a prebunking measure, yet its sustainable impact remains a challenge (Adjin-Tettey, 2021). This paper presents fact-checking based on the principles of peer production (Benkler 2002; 2006) as an approach that integrates and reinforces both prebunking and debunking. It outlines the theoretical foundation and an application example: the CORRECTIV.Faktenforum.

About Peer Production

For CORRECTIV.Faktenforum the non-profit media house Correctiv developed an online platform that enables collaboration and low-threshold participation in fact-checking for volunteers without

professional journalistic training, while allowing to safeguard professional standards. Various models have been applied to conceptualize and analyze such forms of participation in journalism (Abbott 2017). In contrast to crowd-sourcing (Draws et al., 2022) and citizen journalism (Min et al., 2025), the concept of peer production provides a more fitting theoretical lens. Peer production emphasizes not only knowledge creation but also the sharing of values. Key principles of such communities include open access to resources, participation, transparency, and democratic organization (Benkler, 2002).

"Facilitated by the technical infrastructure of the Internet, the hallmark of this socio-technical system is collaboration among large groups of individuals who cooperate effectively to provide information, knowledge, or cultural goods without relying on market pricing" (Benkler, 2006). Originally applied to open-source software development, peer production has also been utilized to study the journalistic initiatives such as WikiTribune (O'Riordan, 2020), which adapted Wikipedia's collaborative model for evidence-based journalism.

About CORRECTIV.Faktenforum

In the case of CORRECTIV.Faktenforum, peer production results in articles compiled through collaborative efforts in the form of structured fact collections, a specific article structure developed for the project. The objective is not limited to producing debunking content. Participants who consistently engage in fact-checking internalize and reinforce practical research skills as well as journalistic standards–achievements that traditional MILT approaches often fail to deliver effectively.

As Benkler (2006) argues, "socio-technical systems of commons-based peer production offer not only a medium of production for various information goods but serve as a context for positive character formation. (...) They offer an opportunity for more people to engage in practices that permit them to exhibit and experience virtuous behavior." Within CORRECTIV.Faktenforum, participants experience journalistic standards such as objectivity, source balance, and the differentiation between opinion and fact through active engagement.

A Survey on the Faktenforum Community

Benkler's concept has proven useful to theorize the Faktenforum approach but also to study the Faktenforum community itself. Correctiv conducted an online questionnaire with closed as well as open-ended questions to examine the composition and motivation of participants as well as their use of the acquired knowledge. Distributed via the organization's newsletters and social media channels in April 2025, the survey was open for ten days and completed by 790 respondents. Of these, 267 had already participated in Faktenforum activities, and 98 were registered on the platform. The following results are derived from respondents reflecting on their activities in the project,

analyzed along the functional dimensions of peer production: motivation, coordination, and integration (O'Riordan et al., 2020).

Motivation: Previous research has identified diverse drivers of engagement in peer production, categorized into three types of individual-level motivations: intrinsic, internalized extrinsic, and extrinsic (Spaeth & Niederhöfer, 2020). According to the survey results, intrinsic motivations were the most prevalent (60.10%), with ideology being a particularly strong driver (41.97%). Among the group of internalized extrinsic motivations, own-use value was a significant factor (21.76%).

Coordination: Most respondents participated in less complex, more individual activities such as disseminating fact collections (26.59%) and submitting potentially false claims (25.47%). Participation was lower in more demanding tasks such as workshops (16.01%) and collaborative debunking (14.61%).

Integration: To grasp this third challenge, the survey asked about the application of knowledge gathered in CORRECTIV.Faktenforum for subsequent prebunking or debunking activities. To categorize their responses, the definition of Ecker et al. (2022) was useful. A majority of respondents (79.4%) reported that they had learned something that helped them to better understand disinformation or to respond to specific false claims. Most of these instances involved prebunking efforts—such as media literacy interventions or preemptive warnings (10%) before exposure to misinformation. Around 30% of responses indicated debunking efforts, where participants used either community research or their own fact-checking to counter specific false claims they encountered.

Directions for Further Research

Beyond these functional challenges, peer production also involves a second dimension: the opportunity for participants to exhibit and experience virtuous behavior. This aspect, however, cannot be adequately captured through self-assessment in an online survey, which underscores the need for further research. In this regard, CORRECTIV.Faktenforum offers a unique case for longitudinal and comparative studies on anti-disinformation initiatives, unmatched by most other projects in the field.

Disclaimer

Both authors are employed at Correctiv and involved in CORRECTIV.Faktenforum. Faktenforum originated from the academic project consortium noFake funded by the BMBF until February 2025.

References

Abbott, J. Y. (2017). Tensions in the scholarship on participatory journalism and citizen journalism. *Annals of the International Communication Association, 41*(3–4), 278–297. https://doi.org/10.1080/23808985.2017.1387519.

Benkler, Y. (2017). Peer production, the commons, and the future of the firm. *Strategic Organization, 15*(2), 264–274. https://doi.org/10.1177/1476127016652606

Chan, M.-p. S., Jones, C. R., Hall Jamieson, K., & Albarracín, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science, 28*(11), 1531–1546. https://doi.org/10.1177/0956797617714579

Cook, J., Lewandowsky, S., & Ecker, U. K. H. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PLOS ONE, 12*(5), e0175799. https://doi.org/10.1371/journal.pone.0175799

Dame Adjin-Tettey, T. (2022). Combating fake news, disinformation, and misinformation: Experimental evidence for media literacy education. *Cogent Arts & Humanities, 9*(1), 2134946. https://doi.org/10.1080/23311983.2022.2134946

Draws, T., Gadiraju, U., Gadiraju, N., & Houben, G. J. (2022). The effects of crowd worker biases in fact-checking tasks. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 2114–2124). Association for Computing Machinery. https://doi.org/10.1145/3531146.3533106

Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology, 1*(1), 13–29. https://doi.org/10.1038/s44159-021-00006-y

List, J. A., Muir, I., Payne, A. A., & Walker, R. (2024). *Toward an understanding of the economics of misinformation: Evidence from a demand side field experiment on critical thinking* (Working Paper No. 32367). National Bureau of Economic Research. https://doi.org/10.3386/w32367

Martel, C., & Rand, D. G. (2024). Fact-checker warning labels are effective even for those who distrust fact-checkers. *Nature Human Behaviour,* 1–11. https://doi.org/10.1038/s41562-024-01915-3

Min, S. J., Kim, S., & Lee, H. (2025). Citizen journalism: Revisiting the concept and developments. *Journalism, 26*(5), 931–943. https://doi.org/10.1177/14648849231220176

O'Riordan, S., Binns, R., & Liao, T. (2020). Theorizing hybrid models of peer production: A case study of an open collaborative journalism platform. In *Proceedings of the 53rd Hawaii International Conference on System Sciences* (pp. 522–531). University of Hawai'i at Mānoa. https://doi.org/10.24251/HICSS.2020.064

Spaeth, S., & Niederhöfer, S. (2020). User motivations in peer production. In M. O'Neil, C. Pentzold, & S. Toupin (Eds.), *The handbook of peer production* (pp. 123–136). Wiley. https://doi.org/10.1002/9781119537151.ch9

Tay, L. Q., Hurlstone, M. J., Kurz, T., & Ecker, U. K. H. (2022). A comparison of prebunking and debunking interventions for implied versus explicit misinformation. *British Journal of Psychology, 113*(3), 591–607. https://doi.org/10.1111/bjop.12542

Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policymaking.* Council of Europe. First Draft. https://firstdraftnews.org/glossary-items/pdf-wardle-c-derakshan-h-2017-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making-council-of-europe/

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Countering Anti-Democratic Beliefs: Future-Orientation Prebunking Interventions for Cognitive Immunology

Marco Dürr \\ Neu-Ulm University of Applied Sciences \\ Neu-Ulm \\ Germany
marco.duerr@hnu.de

Marten Risius \\ Neu-Ulm University of Applied Sciences \\ Neu-Ulm \\ Germany
marten.risius@hnu.de

Winnifred Louis \\ The University of Queensland \\ Brisbane \\ Australia
w.louis@psy.uq.edu.au

KEYWORDS
anti-democratic; future orientation; cognitive immunology; prebunking intervention

Introduction

Threatening anti-democratic views are pervading western societies (Ecker et al., 2024). In this project we aim to adapt prebunking and attitude inoculation interventions from the counter-disinformation context (Basol, Roozenbeek & Van Der Linden, 2020; Tay et al., 2022) to help build cognitive immunology (Compton, 2013; Norman, Johnson & Van Der Linden, 2024) through a future-oriented mindset (Strathman et al., 1994) against anti-democratic beliefs. Anti-democratic beliefs gravitate towards supporting and envisioning an authoritarian government or regime (Jost, 2024), endorsing society submissively following orders, accepting little to no political rights, and approving the concentration of executive power within one person or a small group (Lindstaedt, 2024). The widespread adoption of anti-democratic beliefs poses the risk of shifting states towards dictatorships, repressing freedom of its citizens.

Cognitive Immunology and Future Orientation

Being confronted with this situation, we need to understand how to prevent the formation of anti-democratic beliefs and how adopted attitudes can be mitigated. Therefore, we aim to develop user-centric online interventions countering anti-democratic views by building cognitive immunology. Cognitive immunology is a process, which functions analogously to the traditional immune system of the body. In doing so, psychological defense mechanisms protect the human mind of threats. They represent beliefs, which are not aligned with the predominant, individual belief system (Norman, Johnson & Van Der Linden, 2024). Therefore, within a cognitive immunology perspective, anti-democratic views represent the threat to a democratic individual.

If the cognitive immune system detects a threat like anti-democratic views, mechanisms like doubt production, are evoked to flag intruding beliefs as malicious. Subsequently, cognitive antibodies, taking the form of counterarguments are formed, neutralizing the threat and protecting the identity of individuals (McGuire, 1961; Norman, Johnson, & Van Der Linden, 2024). Therefore, cognitive immunology is a powerful resource for protecting individuals' minds, against cognitive threats (Norman, Johnson, & Van Der Linden, 2024). This protection occurs within an intervention of attitude inoculation. This intervention technique leverages cognitive immunology processes to alter attitudes (e.g., anti-democratic) through exposing individuals to small doses of these beliefs (Compton, 2013).

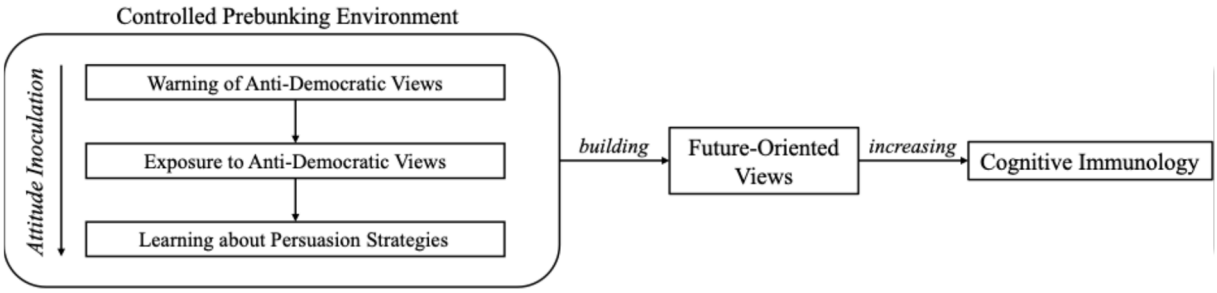
We expect future orientation to be a powerful mean for attitude inoculation against anti-democratic views. Future orientation is a tendency to direct considerations towards distant outcomes of current behaviors. Adopting this perspective results in individuals preferring outcomes, which are attainable within the future compared to outcomes obtainable within the present (Joireman et al., 2012; Strathman et al., 1994). A future-oriented individual will therefore adopt a goal pursuit strategy, which is directed towards pursuing ideal, positive outcomes with an unknown certainty of actual achievement. In contrast, a present-oriented individual will direct their efforts towards avoiding negative consequences (Higgins, 1997; Higgins et al., 2001).

Interventions to Build Cognitive Immunology

We follow the long-term aim of minimizing the impact of anti-democratic attitudes on individuals. Specifically, we aim to adapt attitude inoculation through prebunking interventions, which have been proven to be effective for building cognitive immunology within the domain of disinformation (Basol, Roozenbeek & Van Der Linden, 2020; Tay et al., 2022). We will adapt these interventions to build future-oriented belief systems. Apart from the theoretical considerations above, future orientation is established as a protective resource against anti-social tendencies (Clinkinbeard, 2014; Petrich & Sullivan, 2020). Additionally, authoritarianism is accompanied by more anti-social behaviors, less commitment to values like freedom and moral principles as well as support of anti-democratic people and practices (Jost, 2024). Therefore, developing future-oriented belief systems is the

central building block of effective interventions for developing cognitive immunology against anti-democratic beliefs.

Figure 1: Wireframe representation of the planned intervention strategy and outcome



Apart from protective effects of future orientation, prebunking interventions are an effective tool for strengthening the cognitive immune system within individuals through attitude inoculation. This treatment results in the production of counterarguments, effectively refuting malicious attitudes and beliefs by the treated person (Compton & Pfau, 2005; McGuire, 1964). Attitude inoculation has proven to be an effective instrument counteracting malicious information (Banas & Rains, 2010). Prebunking is a specific form of attitude inoculation. Within such interventions, participants are warned of an incoming threat towards their existing belief system. Afterwards, a small dose of malicious belief is presented alongside a refutation, illustrating the reason for the maliciousness and the persuasion strategies, which simultaneously occur (Tay et al., 2022). As elaborated before, we plan to adapt existing prebunking interventions to the domain of building cognitive immunology against anti-democratic beliefs as depicted in Figure 1. Through this intervention, we additionally aim to build resilience against opinion manipulation while increasing information literacy and critical thinking skills to help protect democratic societies.

Conclusion

Anti-democratic attitudes pose an imminent threat to societies. Interventions focused on building future-oriented belief systems, allow for the protection of individuals by establishing cognitive immunology. By adapting pre-bunking interventions, this immunology is further enhanced through building additional competencies within individuals.

Acknowledgements

This work receives generous funding support from the Bavarian State Ministry of Science and the Arts through the Distinguished Professorship Program as part of the Bavarian High-Tech Agenda.

References

Banas, J. A., & Rains, S. A. (2010). A meta-analysis of research on inoculation theory. *Communication Monographs, 77*(3), 281–311. https://doi.org/10.1080/03637751003704036

Basol, M., Roozenbeek, J., & Van Der Linden, S. (2020). Good news about bad news: Gamified inoculation boosts confidence and cognitive immunity against fake news. *Journal of Cognition, 3*(1), 1–9. https://doi.org/10.5334/joc.91

Clinkinbeard, S. S. (2014). What lies ahead: An exploration of future orientation, self-control, and delinquency. *Criminal Justice Review, 39*(1), 19–36. https://doi.org/10.1177/0734016813501194

Compton, J. (2013). Inoculation theory. In J. Dillard & L. Shen (Eds.), *The SAGE handbook of persuasion: Developments in theory and practice* (pp. 220–236). SAGE Publications.

Compton, J. A., & Pfau, M. (2005). Inoculation theory of resistance to influence at maturity: Recent progress in theory development and application and suggestions for future research. *Annals of the International Communication Association, 29*(1), 97–146. https://doi.org/10.1080/23808985.2005.11679045

Ecker, U., Roozenbeek, J., Van Der Linden, S., Tay, L. Q., Cook, J., Oreskes, N., & Lewandowsky, S. (2024). Misinformation poses a bigger threat to democracy than you might think. *Nature, 630*, 29–32. https://doi.org/10.1038/d41586-024-00620-2

Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist, 52*(12), 1280–1300. https://doi.org/10.1037/0003-066X.52.12.1280

Higgins, E. T., Friedman, R. S., Harlow, R. E., Idson, L. C., Ayduk, O. N., & Taylor, A. (2001). Achievement orientations from subjective histories of success: Promotion pride versus prevention pride. *European Journal of Social Psychology, 31*(1), 3–23. https://doi.org/10.1002/ejsp.27

Joireman, J., Shaffer, M. J., Balliet, D., & Strathman, A. (2012). Promotion orientation explains why future-oriented people exercise and eat healthy: Evidence from the Two-Factor Consideration of Future Consequences-14 scale. *Personality and Social Psychology Bulletin, 38*(10), 1272–1287. https://doi.org/10.1177/0146167212447541

Jost, J. T. (2024). Both-sideology endangers democracy and social science. *Journal of Social Issues, 80*(3), 1138–1203. https://doi.org/10.1111/josi.12606

Lindstaedt, N. (2024). Authoritarianism. In *Encyclopedia Britannica*. https://www.britannica.com/topic/authoritarianism

McGuire, W. J. (1961). The effectiveness of supportive and refutational defenses in immunizing and restoring beliefs against persuasion. *Sociometry, 24*(2), 184–197. https://doi.org/10.2307/2786067

McGuire, W. J. (1964). Some contemporary approaches. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 1, pp. 191–229). Elsevier. https://doi.org/10.1016/S0065-2601(08)60052-0

Norman, A., Johnson, L., & Van Der Linden, S. (2024). Do minds have immune systems? *Journal of Theoretical and Philosophical Psychology*. https://doi.org/10.1037/teo0000223

Petrich, D. M., & Sullivan, C. J. (2020). Does future orientation moderate the relationship between impulse control and offending? Insights from a sample of serious young offenders. *Youth Violence and Juvenile Justice, 18*(2), 156–178. https://doi.org/10.1177/1541204019852611

Strathman, A., Gleicher, F., Boninger, D. S., & Edwards, C. S. (1994). The consideration of future consequences: Weighing immediate and distant outcomes of behavior. *Journal of Personality and Social Psychology, 66*(4), 742–752. https://doi.org/10.1037/0022-3514.66.4.742

Tay, L. Q., Hurlstone, M. J., Kurz, T., & Ecker, U. K. H. (2022). A comparison of prebunking and debunking interventions for implied versus explicit misinformation. *British Journal of Psychology, 113*(3), 591–607. https://doi.org/10.1111/bjop.12545

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Trolling the Trolls

Fighting Dark Participation with Digital Vigilantism

Sabine Barthold \\ TU Dresden | CODIP \\ Dresden \\ Germany
sabine.barthold@tu-dresden.de

Richard Joos \\ TU Dresden | CODIP \\ Dresden \\ Germany
richard.joos@tu-dresden.de

KEYWORDS
dark participation; counter-swarming; cyber vigilantism; swarm culture; community empowerment

Introduction

In this paper, we demonstrate how counter-swarms can act as a form of cyber-vigilantism against dark participation and social injustices using strategies of weaponizing visibility (Trottier, 2017). We argue that communities can be empowered to effectively defend online spaces against enshittification by turning the tactics of trolls and haters against themselves and beating them in their own game, a form of online jiu jitsu, if you will. Numerous instances on social media have demonstrated that individuals engaged in dark participation tend to withdraw from debates when they lose the shield of anonymity and face accountability for their offenses (Munger 2017)—ideally in front of a vast online audience (Polak & Trottier, 2020). While this form of activism, employing counter-swarming tactics, may appear radical, it serves as a means of self-defense. In an increasingly toxic atmosphere surrounding online discussions and attacks against individuals by swarms of trolls, it has allowed victims and vulnerable communities to push back.

Dark Participation in Online Spaces

In January 2025, Meta declared its intention to eliminate fact-checking and content filtering on its primary social media platforms, Instagram and Facebook, to “reaffirm [its] fundamental commitment to free expression” (Meta, 2025). Following Elon Musk's acquisition of Twitter in 2022, his initial action was to dismiss the teams responsible for content moderation and reinstate accounts that

had previously been banned for disseminating racist, transphobic, or violence-glorifying content, as well as conspiracy theories, all under the guise of "protecting freedom of speech." The outcome in both instances was a significant rise in hate speech and disinformation, accompanied by a deterioration of online discourse, a phenomenon that Cory Doctorow (2023) termed "enshittification". This social dynamic, characterized by a blend of despair and rage interspersed with irony and sarcasm, serves as an ideal environment for swarms of trolls, haters, and shit-posters engaged in all forms of dark participation.

Dark participation (Lutz & Hoffmann 2017) seeks to undermine constructive discourse and progressively convert online spaces into toxic and hostile environments, particularly affecting marginalized and vulnerable populations. It has also become a tactic in culture wars (Rao 2018) for digital spaces and is systematically employed by alt-right and reactionary movements to silence people and create hostile spaces for political opponents (e.g. "woke" leftists, climate activists, feminists), stigmatized and marginalized groups ("girls, gays, and theys"), and racial and ethnic minorities. This includes orchestrated shitstorms initiated by influential actors and accounts that pursue their victims across various platforms, often those with little to no moderation, such as Twitch, Reddit, or Telegram. The tactic involves stirring up a sense of persecution, an "us" vs "them" dynamic and the subsequent mobilization of swarms of trolls and haters emboldened by the safety of anonymity. What tactics can effectively counter this development and empower communities within these quasi-public third spaces?

Counter-Swarms: Cyber-Vigilantism as Civic Engagement

Strategies to counter dark participation have focused mainly on two approaches: (1) platform governance, that is, moderation, reporting community guideline violations (e.g., hate speech, disinformation), or flagging or removing harmful content or accounts. (2) Individual defense mechanisms: options to report offensive or criminal behavior to platforms or the police (e.g., harassment, insult, incitement to hatred). These strategies can be effective if online platforms commit to sanctioning destructive user behavior (and put sufficient resources into their enforcement) and if users can rely on law enforcement to effectively prosecute criminal offenses committed online. However, often, both strategies failed. A more performative and collaborative approach examined in this research is the strategy of (3) counter-swarming, a practice that is helping empower communities to combat dark participation.

Communities have started to use swarming as a novel method of political advocacy. Online swarms are networks of interconnected individuals, content, and bots, enabling members to act autonomously while achieving a collective effect through a common focus that is reinforced by the algorithmic feedback loops of their online environments (Fernández, 2023). Generation Z, especially, leverages social media to highlight contemporary political issues and spark conversations around them. Notably, TikTok has become a key platform for young people to voice their political opinions

and participate in political advocacy (Serdar, 2020). In June 2020, a swarm of K-Pop fans and Teens organized on TikTok to troll the Trump election campaign by buying up thousands of tickets for a rally held in Tulsa, Oklahoma only then not to show up leaving the venue largely empty (Lorenz et al. 2020). Swarm culture is reflected in memes, internet challenges or streaming hits, but it can also transgress into offline spaces as happened in fluid and impulsive political movements like #BLM or #MeToo.

Digital vigilantism and counter-swarming attacks have demonstrated their potential as effective mechanisms of community-based self-defense in online spaces, particularly when other modes of protection fail to address immediate threats and harassment (Crockett, 2017). Counter-swarms exploit the same tools and strategies used by trolls—virality, anonymity, and collective action—but invert their purpose to protect rather than harm. This digital vigilantism is part of a broader 'vengeance culture' in the cybersphere: The extrajudicial "punishment" meted out by cyber-vigilantes usually involves some combination of trickery, persuasion, and public shaming (Jane, 2016). Vigilantes are using strategies of weaponizing visibility so that individuals engaged in dark participation withdraw from debates when they lose the shield of anonymity and face accountability for their offenses. It serves as a means of self-defense and has allowed victims and vulnerable communities to push back and reclaim online spaces. This strategy instills a sense of empowerment and agency for victims and their wider communities and a sense of solidarity with other targets of harassment.

However, this strategy comes with certain risks and limitations regarding the logistics, efficacy, and ethics of online vigilantism and swarm dynamics. For instance, identifying and exposing perpetrators may not always be possible, especially when they are using anonymous or "burner" accounts for their attacks. And some of them might even enjoy the publicity, so that exposing their harassment and attacks might be counterproductive and considered "feeding the trolls". Furthermore, activists engaged in digital vigilantism might risk legal repercussions, harassment from opposing groups, or becoming targets of (dangerous) retaliatory attacks (Jane, 2016). Lastly, there are ethical issues and risks to counter-swarming: cyber-swarms, once incited, are hard to control and can cause unpredictable damage, so that the outcomes are sometimes adversarial to intent or reactions might be disproportionate or directed at innocents. The "crowdsourcing" of justice can be unpredictable and puts the responsibility of action on the individual victim instead of platforms and institutions.

Conclusion

Digital vigilantism and counter-swarming attacks have demonstrated their potential as effective mechanisms of community-based self-defense in online spaces, particularly when other modes of protection fail to address immediate threats and harassment. However, any act of digital vigilantism is still likely to have uncertain results, can be ethically questionable, can put activists at risk, and ultimately strengthens extrajudicial cultures online. Thus, the rise of counter-swarms and digital vigilantism reflects a reaction to platform enshittification—the degradation of online spaces into toxic arenas. So, even though cyber vigilantism might bring short-term wins, it often fails to address

the shortcomings of legal and corporate frameworks as well as underlying systematic forms of oppression that communities must face in online spaces. Ultimately, the problems of cyber-harassment and cyber-hate urgently require institutional remedies. Thus, rather than being viewed as definitive solutions, these self-defense strategies should be understood as symptomatic responses to a broader systemic failure—one where corporations, lawmakers, and enforcement agencies have consistently neglected their responsibilities in establishing robust legislative frameworks, implementing effective corporate policies, developing comprehensive pedagogical approaches, and creating sustainable social structures necessary for addressing these issues at their fundamental institutional levels. The complex nature of cyber-harassment and cyber-hate demands sophisticated institutional remedies that go beyond individual interventions. This underscores the critical importance of implementing systemic changes and establishing comprehensive legal frameworks to effectively combat these issues. Well-designed institutional approaches have the potential to deliver more sustainable, equitable, and ethically sound solutions to online harassment and hate, particularly when compared to the limitations and risks associated with individual acts of vigilantism.

References

Crockett, M. J. (2017). Moral outrage in the digital age. *Nature Human Behaviour, 1*(11), 769–771. https://doi.org/10.1038/s41562-017-0213-3

Doctorow, C. (2023, January). Commentary: Cory Doctorow: Social quitting. *Locus Online*. https://locusmag.com/2023/01/commentary-cory-doctorow-social-quitting/

Fernández, R. (2023). Welcome to the swarm. *Summer of Protocols*. https://summerofprotocols.com/the-swarm-and-the-formation-web

Jane, E. A. (2016). Online misogyny and feminist digilantism. *Continuum, 30*(3), 284–297. https://doi.org/10.1080/10304312.2016.1166560

Lorenz, T., Browning, K., & Frenkel, S. (2020, June 21). TikTok teens and K-pop stans say they sank Trump rally. *The New York Times*. https://www.nytimes.com/2020/06/21/style/tiktok-trump-rally-tulsa.html

Lutz, C., & Hoffmann, C. P. (2017). The dark side of online participation: Exploring non-, passive, and negative participation. *Information, Communication & Society, 20*(6), 876–897. https://doi.org/10.1080/1369118X.2017.1293129

Meta. (2025, January 7). More speech and fewer mistakes. *Meta Newsroom*. https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/

Munger, K. (2017). Tweetment effects on the tweeted: Experimentally reducing racist harassment. *Political Behavior, 39*(3), 629–649. https://doi.org/10.1007/s11109-016-9373-5

Polak, S., & Trottier, D. (2020). Introducing online vitriol. In *Violence and trolling on social media: History, affect, and effects of online vitriol* (pp. 9–20). Amsterdam University Press.

Rao, V. (2018, March 6). A quick (battle) field guide to the new culture wars. *Ribbonfarm*. https://www.ribbonfarm.com/2018/03/06/a-quick-battle-field-guide-to-the-new-culture-wars/

Serdar, K. (2020, July 23). Opinion: Teens are finding power through TikTok. *HS Insider | LA Times*. https://highschool.latimes.com/la-canada-high-school/opinion-teens-are-finding-power-through-tiktok/

Trottier, D. (2017). Digital vigilantism as weaponisation of visibility. *Philosophy & Technology, 30*(1), 55–72. https://doi.org/10.1007/s13347-016-0216-4

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Ethical Foundations for Resilient Smart Cities

Preliminary Insights from Stakeholder Perspectives on Crisis Communication Technologies

Yannick Fernholz \\ Weizenbaum Institute \\ Berlin \\ Germany
yannick.fernholz@weizenbaum-institut.de

KEYWORDS
smart cities; ethical design; value-sensitive design

Abstract

The concept of smart cities has emerged as a response to the growing challenges of urbanization, aiming to leverage Information and Communication Technologies (ICT) and the Internet of Things (IoT) to improve the quality of life for citizens (Del-Real et al., 2023). By optimizing urban services, fostering sustainability, and enhancing governance, smart cities promise significant societal benefits. However, these technological advancements are not without criticism. Scholars have pointed out the risk of top-down, technocratic implementations that prioritize efficiency and corporate interests over inclusivity and the actual needs of citizens (Emami-Naeini et al., 2019; Anthopoulos, 2017; Fernandez-Anez et al., 2018). Moreover, ethical concerns, such as issues of privacy, equity, and the digital divide, remain insufficiently addressed, particularly in scenarios requiring rapid decision-making, such as crisis management. This study investigates how ethical principles can guide the design of smart city technologies, with a specific focus on crisis communication as a use case, while situating the discussion within the broader context of smart city development.

This research adopts a Design Science Research (DSR) approach (Hevner, 2010) and integrates the Value Sensitive Design (VSD) methodology (Friedman et al., 2013), which emphasizes the alignment of technological design with human values. Additionally, the Penta Helix model (Satyam & Calzada, 2017) serves as a guiding framework to include diverse perspectives from government, academia, the private sector, civil society, and activists. This multi-stakeholder approach recognizes that the success of smart city projects depends on collaboration between these groups to address conflicting priorities and trade-offs effectively. Through semi-structured expert interviews, we explore how

stakeholders define smart cities, prioritize ethical considerations, and perceive barriers to the implementation of value-sensitive technologies.

Preliminary findings from the interviews reveal that stakeholders exhibit significantly divergent perspectives on the role and purpose of smart city technologies. Government officials and private sector actors tend to focus on efficiency, scalability, and technological innovation as key drivers of smart city projects. These priorities align with traditional governance models that emphasize centralized control and rapid decision-making capabilities. In contrast, representatives from civil society and activists emphasize the importance of inclusivity, equity, and community-driven solutions. For example, several interview participants highlighted that the design of smart city technologies for crisis communication often overlooks the needs of marginalized groups, exacerbating the digital divide and creating barriers to accessing critical services. These insights resonate with broader critiques in the literature, which argue that smart cities risk reproducing existing social inequalities if inclusivity is not made a central design principle (Macke et al., 2018; Fernandez-Anez et al., 2018). Another key theme emerging from the interviews is the tension between the need for centralized data systems and the ethical implications of such systems. Centralized data platforms, while enabling more efficient resource allocation and crisis management, also raise significant concerns about privacy and surveillance. Stakeholders frequently cited fears of data misuse, lack of transparency, and inadequate regulatory frameworks to safeguard citizen rights. These findings echo existing studies that highlight the need to balance privacy and security in smart city environments (Kitchin, 2019; Emami-Naeini et al., 2019). Stakeholders expressed a desire for transparent data governance structures that ensure accountability while protecting individual privacy.

A further insight from the interviews concerns the operationalization of ethical principles. While high-level values such as sustainability, equity, and transparency are widely acknowledged as desirable, their practical implementation often proves challenging. Many stakeholders noted that existing guidelines remain abstract and lack actionable steps for integrating ethical considerations into design processes. This gap is particularly evident in crisis scenarios, where the urgency of decision-making often leads to ethical trade-offs that are neither well-documented nor sufficiently debated. As Vidiasova et al. (2019) argue, operationalizing ethics in smart cities requires frameworks that are not only theoretically robust but also practically feasible. These preliminary findings highlight the complex and often conflicting priorities of stakeholders involved in smart city development. By examining these tensions through the lens of crisis communication, the study provides initial insights into how value-sensitive design approaches can be used to navigate ethical trade-offs and foster the development of inclusive and equitable technologies. The study contributes to the emerging field of ethical smart city research by offering empirically grounded insights into stakeholder priorities, value conflicts, and barriers to implementation. Furthermore, it seeks to bridge the gap between theoretical frameworks and practical applications, providing actionable guidance for policymakers, technologists, and urban planners working to create resilient, inclusive, and ethically sound urban environments.

References

Anthopoulos, L. G. (2017). *Understanding smart cities: A tool for smart government or an industrial trick?* (Vol. 22). Springer International Publishing. https://doi.org/10.1007/978-3-319-57015-0

Del-Real, C., Ward, C., & Sartipi, M. (2023). What do people want in a smart city? Exploring the stakeholders' opinions, priorities and perceived barriers in a medium-sized city in the United States. *International Journal of Urban Sciences, 27*(sup1), 50–74. https://doi.org/10.1080/12265934.2023.2189352

Emami-Naeini, P., Dixon, H., Agarwal, Y., & Cranor, L. F. (2019). Exploring how privacy and security factor into IoT device purchase behavior. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–12). ACM. https://doi.org/10.1145/3290605.3300764

Fernandez-Anez, V., Fernández-Güell, J. M., & Giffinger, R. (2018). Smart City implementation and discourses: An integrated conceptual model. The case of Vienna. *Cities, 78*, 4–16. https://doi.org/10.1016/j.cities.2017.12.004

Friedman, B., Kahn, P. H., Borning, A., & Huldtgren, A. (2013). Value sensitive design and information systems. In N. Doorn, D. Schuurbiers, I. van de Poel, & M. E. Gorman (Eds.), *Early engagement and new technologies: Opening up the laboratory* (pp. 55–95). Springer. https://doi.org/10.1007/978-94-007-7844-3_4

Hevner, A., & Chatterjee, S. (2010). *Design research in information systems: Theory and practice* (Vol. 22). Springer Science & Business Media. https://doi.org/10.1007/978-1-4419-5653-8

Kitchin, R. (2019). Toward a genuinely humanizing smart urbanism. In F. Cardullo, C. Di Feliciantonio, & R. Kitchin (Eds.), *The right to the smart city* (pp. 193–204). Emerald Publishing Limited. https://doi.org/10.1108/9781787691391-014

Macke, J., Casagrande, R. M., Sarate, J. A. R., & Silva, K. A. (2018). Smart city and quality of life: Citizens' perception in a Brazilian case study. *Journal of Cleaner Production, 182*, 717–726. https://doi.org/10.1016/j.jclepro.2018.02.078

Satyam, A., & Calzada, I. (2017). *The smart city transformations: The revolution of the 21st century*. Bloomsbury Publishing.

Vidiasova, L., Cronemberger, F., & Vidiasov, E. (2019). Risk factors in smart city development in Russia: A survey. In A. Chugunov, I. Magdalinou, Y. Misnikov, & N. Trutnev (Eds.), *Electronic Governance and Open Society: Challenges in Eurasia: 5th International Conference, EGOSE 2018*, St. Petersburg, Russia, November 14–16, 2018, Revised Selected Papers (Vol. 5, pp. 26–37). Springer International Publishing. https://doi.org/10.1007/978-3-030-39815-6_3

Del-Real, Cristina, Chandra Ward, and Mina Sartipi. What do people want in a smart city? Exploring the stakeholders' opinions, priorities and perceived barriers in a medium-sized city in the United States. *International Journal of Urban Sciences 27*, no. sup1 (2023): 50-74.

Emami-Naeini, Pardis, Henry Dixon, Yuvraj Agarwal, and Lorrie Faith Cranor. Exploring how privacy and security factor into IoT device purchase behavior. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1-12. 2019.

Fernandez-Anez, Victoria, José Miguel Fernández-Güell, and Rudolf Giffinger. Smart City implementation and discourses: An integrated conceptual model. The case of Vienna. *Cities 78* (2018): 4-16.

Friedman, Batya, Peter H. Kahn, Alan Borning, and Alina Huldtgren. Value sensitive design and information systems. *Early engagement and new technologies: Opening up the laboratory* (2013): 55-95.

Hevner, Alan, and Samir Chatterjee. *Design research in information systems: theory and practice*. Vol. 22. Springer Science & Business Media, 2010.

Kitchin, Rob. Toward a genuinely humanizing smart urbanism. In *The right to the smart city*, pp. 193-204. Emerald Publishing Limited, 2019.

Macke, Janaina, Rodrigo M. Casagrande, João Alberto R. Sarate, and Kelin A. Silva. Smart city and quality of life: Citizens' perception in a Brazilian case study. *Journal of cleaner production 182* (2018): 717-726.

Satyam, Amitabh, and Igor Calzada. *The smart city transformations: The revolution of the 21st century*. Bloomsbury Publishing, 2017.

Vidiasova, Lyudmila, Felippe Cronemberger, and Evgenii Vidiasov. Risk factors in smart city development in Russia: A survey. In *Electronic Governance and Open Society: Challenges in Eurasia: 5th International Conference*, EGOSE 2018, St. Petersburg, Russia, November 14-16, 2018, Revised Selected Papers 5, pp. 26-37. Springer International Publishing, 2019.

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Panel 4

Session 10: Infrastructure and Market Dynamics

Session 11: Data Disputes

Session 12: Democratic Wellbeing: Democracy and Media in Europe

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Regulation of Digital Platforms Work and the UN's 2030 Agenda

How to Establish Compatibility Between Objectives and Approaches?

Felipe Mano \\ São Paulo State University \\ Franca \\ Brazil
felipe.mano@unesp.br

KEYWORDS
legal regulation; platform work; UN's 2030 agenda

Introduction

The growth of digital work platforms has challenged traditional legal frameworks, revealing limitations in state regulation in the face of the complexity of these business models. This research analyzes the compatibility of current legislation and draft laws with the ethical guidelines of the UN's 2030 Agenda, especially its SDG 8, which promotes decent work and inclusive economic growth. The hypothesis is that, although some legal advances exist, most regulatory approaches ignore the structural foundations of precariousness, especially financialization and data extraction, limiting their effectiveness in transforming working conditions on digital platforms.

Objectives

The overall objective of the research was to analyze the compatibility of legal regulatory measures on platform work with the ethical guidelines of the UN 2030 Agenda, in particular its SDG 8, considering whether and how such rules address the business models of digital platforms.

The specific objectives were: a) to map laws on platform work; b) to examine the guidelines of the 2030 Agenda and the objectives of SDG 8; c) to assess the compatibility of laws with SDG 8; and d) to analyze the legal treatment given to business models.

Methodology

The research adopted a qualitative approach, focusing on the documentary analysis of legislation on platform work. Data collection was carried out through desk research, using as its main source the report *Realizing Decent Work in the Platform Economy*, published by the International Labor Organization (ILO) in January 2024. The document brings together a broad set of legal frameworks adopted by different countries up to the date of publication (whether in vigor or not) and presents different criteria for analysis, including the treatment of employment relationships, social protection, and platform responsibility.

The data was analyzed inductively, with an emphasis on verifying the compatibility of these laws with the ethical guidelines of the UN's 2030 Agenda and its SDG 8, which deals with the promotion of decent work and economic growth. It also examined whether and how the business models of digital platforms are considered in legal texts, as a way of assessing the scope and limitations of the legal approaches analyzed. From this basis, a deductive approach was used, articulating empirical findings with critical references on digital work and law. The aim was to reflect on the viability and contours of regulatory alternatives more consistent with the principles and objectives of the 2030 Agenda.

Development

Law is a central instrument in the regulation of social relations, especially in the economic field, which includes labor relations. Although there are complementary forms of regulation, such as market self-regulation and co-regulation schemes between states and private actors, legal regulation plays a structuring role in the way society is organized (Keller, 2019).

However, its actions tend to reproduce forms of sociability linked to the capitalist mode of production, consolidating certain imaginaries of economic and labor organization (Jasanoff, 2015; Klare, 1977, 2002; Pachukanis, 2017). These forms follow historical regimes of accumulation and predominant modes of regulation, reflecting normative limits that shape the scope of legal interventions (Klare, 1977, 2002; Lipietz, 1993). Thus, legal frameworks often reinforce existing structures, including in the treatment of digital platform business models.

The report *Realizing Decent Work in the Platform Economy* (ILO, 2024) identified 38 laws on platform work in 20 countries on four continents. The regulations were analyzed according to five criteria: a) types of platforms covered (online and/or location-based); b) recognition of employment relationships; c) social security and labor guarantees; d) social participation in regulatory development; and e) “other areas,” which includes data protection, use of algorithms, and licensing of platforms and workers. Of the total, 15 laws cover online platforms, 37 cover location-based platforms, and 14 cover both. Only 13 recognize employment relationships; 33 ensure social and labor protections; 5 involved social representation in their formulation; and 22 address aspects such as algorithms, data, and licensing.

These measures were compared with the guidelines of SDG 8 of the UN 2030 Agenda, which proposes sustainable economic growth and productive and decent work for all (UN ; [n.d.]). The analysis indicates that, although many laws advance in guaranteeing rights, few address the structures that sustain the precariousness of work on digital platforms, especially business models based on financialization, intensive use of data, algorithmic control, and appropriation of digital infrastructure (Mano; 2024).

The link between financialization and data extraction sustains a logic in which the reproduction of capital is independent of the direct production of goods and services, shifting labor to a marginal position in the productive system. Capital begins to be valued based on its own financial and informational dynamics. In the business models of digital platforms, this correlation between finance, technology, and data extraction is fundamental for capturing income from other sectors of the economy, without necessarily generating value through production. These characteristics define platform capitalism and contribute directly to the precariousness of work, by reconfiguring productive relations and weakening traditional legal and social protections for workers (Grohmann, 2025; Mano, 2024; Paraná, 2024; 2025; Sadowski, 2019).

In this sense, it is argued that overcoming precarious working conditions on digital platforms through legal regulations requires legal frameworks that address these structural issues, going beyond the formal recognition of rights. Effective regulatory approaches need to consider how platforms operate, appropriate value, and organize work, at the risk of promoting only superficial changes. Thus, in order to align with the ethical horizon of SDG 8, such measures must also integrate a critical analysis of business models and their impact on labor relations (Mano, 2024).

Conclusions

Overcoming precariousness in digital platforms requires a multisectoral regulatory policy that addresses business models and their structural foundations. This includes combating data colonialism, protecting workers, strengthening digital sovereignty, providing access to technological infrastructure, and promoting participatory forms of governance. Measures such as the digital solidarity economy, already underway in countries such as Brazil, point to possible paths forward. To align with SDG 8 of the 2030 Agenda, legal frameworks must transcend the recognition of rights and address the logic that sustains the current precariousness, including by articulating different forms of regulatory intervention, such as co-regulation.

References

Grohmann, R. (2025). Free labor and data labor in the digital economy. In T. Venturini, A. Acker, J.-C. Plantin, & T. Walford (Eds.), *The Sage handbook of data and society* (pp. 109–118). Sage Publications.

International Labour Organization (ILO). (2024). *Realizing decent work in the platform economy*. ILO.

Jasanoff, S. (2015). Future imperfect: Science, technology, and the imaginations of modernity. In S. Jasanoff & S.-H. Kim (Eds.), *Dreamscapes of modernity: Sociotechnical imaginaries and the fabrication of power* (pp. 1–33). University of Chicago Press.

Keller, C. I. (2019). *Regulação nacional dos serviços de internet: Exceção, legitimidade e o papel do Estado* [Doctoral thesis, Universidade do Estado do Rio de Janeiro (UERJ)].

Klare, K. (1977). Judicial deradicalization of the Wagner Act and the origins of modern legal consciousness, 1937–1941. *Minnesota Law Review, 62*, 265–339. https://heinonline.org/HOL/LandingPage?handle=hein.journals/mnlr62&div=19&id=&page=

Klare, K. (2002). The horizons of transformative labour and employment law. In J. Conaghan, R. M. Fischl, & K. Klare (Eds.), *Labour law in an era of globalization: Transformative practices & possibilities* (pp. 3–29). Oxford University Press.

Lipietz, A. (1993). From Althusserianism to "regulation theory." In E. A. Kaplan & M. Sprinker (Eds.), *The Althusserian legacy* (pp. 99–138). Verso.

Mano, F. G. (2024). Relaciones de producción, derecho y modelos de negocio: La regulación del trabajo en plataformas. *Nuevo Derecho, 20*(35), 1–34. https://doi.org/10.25057/2500672x.1640

Pachukanis, E. B. (n.d.). *Teoria geral do direito e marxismo* (P. V. d. Almeida, Trans.). Boitempo.

Paraná, E. (2024). Platform studies and the finance-technology nexus: For a "genetic" approach. *Platforms & Society, 1*. https://doi.org/10.1177/29768624241286779

Paraná, E. (2025). Financialised digitalisation, digitalised financialisation: The inseparability between technological domination and financial hegemony in contemporary capitalism. In A. Kangas, I. Gataulina, M. Poutanen, A. I. Rajala, & H.-E. Ventovirta (Eds.), *Retheorising capitalism* (pp. 263–280). Tempere University.

Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data & Society, 6*(1), 205395171882054. https://doi.org/10.1177/2053951718820549

United Nations (UN). (n.d.). "Goal 8". United Nations. SDG. https://sdgs.un.org/goals/goal8

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Introducing a New Methodology: Composition-Based Classification of Online Videos

David Wegmann \\ Aarhus University \\ Aarhus \\ Denmark
david.wegmann@cc.au.dk

Marcus Olesen \\ Aarhus University \\ Aarhus \\ Denmark

Anja Bechmann \\ Aarhus University \\ Aarhus \\ Denmark
anjabechmann@cc.au.dk

KEYWORDS
YouTube; social media video classification; video as data; multimodal clustering; computational communication science; data donation

Introduction

YouTube, a central component of modern digital media, serves as a repository for diverse content—practical instructions, entertainment, education, and political self-expression. However, its user-generated nature raises concerns about the prevalence of misinformation and radicalizing content, compounded by the platform's opaque algorithms and immense content diversity (McGrady et al., 2023; Ribeiro et al., 2020; Tang et al., 2021). These challenges highlight the urgent need for scalable methods to analyze and classify online videos effectively (Hussain et al., 2018). This study introduces a novel, composition-based methodology for classifying online videos, enabling researchers to analyze extensive datasets and address critical questions about content consumption and algorithmic curation.

Objectives

This research aims to develop an automated, bottom-up video classification method capable of categorizing vast amounts of content while maintaining interpretability and transparency. Using a

dataset of 20,000 YouTube videos representing the viewing habits of Danes, the study seeks to identify meaningful clusters of videos, such as educational, entertainment, or radicalizing content. Denmark provides a highly relevant case study, as 50% of its adult population uses YouTube, with 10% relying on it for news (Schrøder et al., 2023).

Methodology

Recent advancements in deep learning-based multimodal learning have refined methods to the point where transfer learning, leveraging vector representations from pre-trained general-purpose models, now matches or surpasses embeddings based on technical content characteristics (Wedel et al., 2024). However, even state-of-the-art video learning models face significant limitations when processing the diversity content as diverse as online videos. Therefore, this study focusses on directly measurable features of video composition, such as linguistic content, visual elements, and audio characteristics. This transparent approach enhances interpretability, allowing researchers to retrace and critically assess classification processes. The foundation of this study is a data donation project involving 1,000 Danish participants who contributed their YouTube watch histories through Google Takeout. After screening for sufficient YouTube usage, 20,000 publicly available videos were sampled from participants' watch histories over the past five years. Ethical considerations, including GDPR compliance and IRB approval, ensured participant anonymity and data security throughout the process.

The videos and their metadata were downloaded using yt-dlp1. Linguistic features were extracted using WhisperAI2 for transcription and TextDescriptives3 for low-level text descriptors (Hansen et al., 2023). Following prior research Visual aspects were analyzed by dividing videos into 10-second segments, extracting keyframes, and calculating parameters such as color distributions and optical flow using FFmpeg4 and OpenCV5 (Berger et al., 2021; Zhu et al., 2024). Audio features were processed with Librosa6, generating segment-wise metrics such as spectral contrast and rhythmic characteristics. These analyses yielded 327 parameters per video, encompassing linguistic, visual, and audio modalities.

1 https://github.com/yt-dlp/yt-dlp

2 https://github.com/openai/whisper

3 https://github.com/HLasse/TextDescriptives

4 https://ffmpeg.org

5 https://github.com/opencv/opencv

6 https://github.com/librosa/librosa

Preliminary Results

A proof-of-concept analysis was conducted on a preliminary dataset of 1,114 videos from 200 participants. The videos had an average duration of 12 minutes, with a mean of 266,490 likes and an average online lifespan of 3.87 years. WhisperAI identified 73.97% of the videos as English, 20.56% as Danish, with confidence levels exceeding 94% on average. Initial clustering revealed five meaningful categories, informed by linguistic and visual features, with significant potential for scalability to the full dataset.

Discussion

This study tests the feasibility of a scalable, automated methodology for classifying online videos, addressing critical gaps in current research. By integrating linguistic, visual, and audio features, the approach offers a structured framework for analyzing video content without reliance on subjective classifications or socially influenced metadata. The preliminary results validate the method's potential for identifying meaningful clusters within large, diverse video datasets. However, limitations of the proof-of-concept analysis, such as the small sample size and reliance on early-stage data, necessitate further validation with the full dataset of 20,000 videos.

Conclusion

This research introduces a novel method for classifying large-scale online video datasets, combining interpretability with scalability. By focusing on computationally extractable features, the approach reduces reliance on opaque algorithms and socially biased metadata, offering a transparent and adaptable framework for computational media research. The preliminary findings underscore the method's potential for advancing our understanding of online video consumption patterns, particularly in the context of Danish media habits.

Ultimately, this methodology provides a flexible tool for analyzing online videos across diverse populations and platforms. Its applicability extends to critical questions about content exposure, algorithmic influence, and the role of digital media in shaping individual and collective experiences. By empowering researchers with scalable and transparent analytical tools, this study contributes to broader efforts to enhance democratic participation and individual well-being in digital societies.

References

Berger, J., Kim, Y. D., & Meyer, R. (2021). What makes content engaging? How emotional dynamics shape success. *Journal of Consumer Research, 48*(2), 235–250. https://doi.org/10.1093/jcr/ucab010

Hansen, L., Olsen, L. R., & Enevoldsen, K. (2023). *TextDescriptives: A Python package for calculating a large variety of metrics from text* (No. arXiv:2301.02057). arXiv. https://doi.org/10.48550/arXiv.2301.02057

Hussain, M. N., Tokdemir, S., Agarwal, N., & Al-Khateeb, S. (2018). Analyzing disinformation and crowd manipulation tactics on YouTube. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 1092–1095). https://doi.org/10.1109/ASONAM.2018.8508766

McGrady, R., Zheng, K., Curran, R., Baumgartner, J., & Zuckerman, E. (2023). Dialing for videos: A random sample of YouTube. *Journal of Quantitative Description: Digital Media, 3*. https://doi.org/10.51685/jqd.2023.022

Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A. F., & Meira, W. (2020). Auditing radicalization pathways on YouTube. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 131–141). https://doi.org/10.1145/3351095.3372879

Schrøder, K., Blach-Ørsten, M., & Eberholst, M. K. (2023). *Danskernes brug af nyhedsmedier 2023* (Rapport No. 978-87-973514-3-7). Center for Nyhedsforskning, Roskilde Universitet. https://doi.org/10.5281/zenodo.7956294

Tang, L., Fujimoto, K., Amith, M. (Tuan), Cunningham, R., Costantini, R. A., York, F., Xiong, G., Boom, J. A., & Tao, C. (2021). "Down the rabbit hole" of vaccine misinformation on YouTube: Network exposure study. *Journal of Medical Internet Research, 23*(1), e23262. https://doi.org/10.2196/23262

Wedel, L., Ohme, J., & Araujo, T. (2024). Augmenting data download packages – Integrating data donations, video metadata, and the multimodal nature of audio-visual content. *Methods, Data, Analyses, 0*, 1–32. https://doi.org/10.12758/mda.2024.08

Zhu, J., Cheng, M., & Wang, Y. (2024). Viewer in-consumption engagement in pro-environmental tourism videos: A video analytics approach. *Journal of Travel Research*. https://doi.org/10.1177/00472875231219634

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Beyond the EU Data Act: Value of IoT Data, Market Failures, and Consumer Choice in the B2C Sector

Wolfgang Kerber \\ University of Bonn \\ Bonn \\ Germany
kerber@wiwi.uni-marburg.de

Karsten K. Zolna \\ University of Marburg \\ Marburg \\ Germany
karsten.zolna@wiwi.uni-marburg.de

KEYWORDS
internet of things; EU Data Act; data governance; data use contracts; consumer protection

Abstract

The recently enacted EU Data Act has introduced significant changes of the governance of IoT data (Internet of Things), which so far is dominated by IoT manufacturers. This includes new rights of users to access and share IoT data as well as a new contract which data holders (manufacturers) need for getting permission to use non-personal IoT data from the users. This article analyzes from an economic market failure perspective the problems that consumers face with regard to getting meaningful control over the generation, collection, use, and sharing of their non-personal and personal IoT data. It is shown that, in B2C contexts, serious market failures exist with regard to the terms of the "licensing" contracts of personal and non-personal IoT consumer data (buy-out contracts, adverse selection), which can lead to a too large collection and sharing of IoT data as well as to not enough choice for the consumers. After presenting a broad overview about the current policy discussion about a minimum of choice for consumers in competition law and data protection law (German Facebook case, new Opinion of EDPB about "consent or pay" on online platforms, Art. 5(2) DMA and opt-out solutions in California privacy law), the last part of the article explores and analyzes a wide range of different policy options for granting consumers more choice about the collection, use, and sharing of their IoT data (consumer empowerment). One of these policy solutions would give consumers the right to use their IoT devices also without having to "license" to data hold

ers IoT data that are not necessary for the functionality of these devices, which also might lead to "consent or pay" solutions in IoT governance. Other policy solutions would give consumers more direct de facto control over their IoT devices and their IoT data, which also might require more technical interoperability and standardisation.

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

It's Complicated: DSA-Based Data Access and the Relationship Status of Platforms, Regulators, and Researchers

LK Seiling \\ DSA40 Data Access Collaboratory \\ Berlin \\ Germany
lukas.seiling@dsa40collaboratory.eu

Jakob Ohme \\ Weizenbaum Institute \\ Berlin \\ Germany
jakob.ohme@weizenbaum-institut.de

Ulrike Klinger \\ University of Amsterdam \\ Amsterdam \\ Netherlands
u.klinger@uva.nl

KEYWORDS
Digital Services Act; Data Access; risk governance; evidence-based decision-making; platform regulation

The Digital Services Act, Systemic Risk, and Data Access

With the Digital Services Act (DSA), adopted in 2022 and fully applicable since February 2024 (European Parliament and Council of the European Union, 2022), the European Commission (EC) can designate intermediary services with more than 45 million monthly service recipients in the EU as Very Large Online Platforms or Search Engines (VLOPSEs). This threshold means that platforms which control information flow to and between a significant share of EU citizens, are required to follow strict risk management obligations. Specifically, VLOPSEs are required to assess and mitigate "any systemic risks in the Union stemming from the design or functioning of their service and its related systems [...], or from the use made of their services." (Art. 34(1), Art. 35 DSA). They also have the duty to disclose information through a variety of processes (Leerssen, 2024), among them to provide vetted researchers with access to data for "the sole purpose of conducting research that contributes to the detection, identification and understanding of systemic risks [...], and to the assessment of the adequacy, efficiency and impacts of the risk mitigation measures" (Art. 40(4) DSA). This work will not attempt to resolve the paradoxes arising from the usage of the term "systemic

risk", a concept without clear theoretical or disciplinary bounds (Centeno et al., 2015; De Bandt & Hartmann, 2000; Helbing, 2013) in the context of a purpose limitation on data use. Instead, it will attempt to provide a brief characterisation and diagnosis of the now formalised relationship between platforms, researchers, and regulators, meant to facilitate research on systemic risks through data access.

An Imbalanced Triad with Unclear Boundaries

While the national enforcement authorities (Digital Service Coordinators, DSCs) are tasked with parts of the data access authorisation procedure (Art. 40(4-11)), the DSA grants exclusive supervisory and enforcement authority for the specific obligations for VLOPSEs to the EC (Buri & van Hoboken, 2022, Art. 56(2) DSA). Provided that it does not capture the entirety of stakeholders in the DSA's risk management regime (Griffin, 2025a) this triadic constellation between researchers, platforms, and enforcement authorities is central to the development of this governance approach (Goanta et al., 2025). The knowledge production process around systemic risk set out by Art. 40 drives the DSA's theory of change. Thus for the legislation to succeed at creating "a safe, predictable and trusted online environment" (Art. 1(1) DSA), consensus and cooperation must be achieved between actors with widely divergent motivations, access to resources and information, network position, and potential action space.

For researchers, the DSA's data access provisions present both an unseen potential to better understand large online platforms (Valkenburg et al., 2024) as well as a significant limitation of scientific freedom due to its purpose limitation (Mast, 2024). At the same time, research resulting from DSA-based data access will feed back into the legislation's specific implementation by informing non-compliance procedures or political agenda setting (Leerssen, 2024), begging researchers to reflect on the role they should play in democratic risk governance – especially given that the EC provides little transparency on the DSA enforcement strategy (Fabbri, 2025). With increased transparency and observability (Rieder & Hofmann, 2020), platforms, on the other hand, face potential enforcement action by the authorities as well as considerable compliance, reputation, product adaptation, and competitive disadvantage costs.

This imbalanced setup is exacerbated by its unequal distribution of responsibilities: while data can mostly be exclusively accessed on the platforms' terms[1], the burden to both practically (through the establishment of standard measures and practices for data verification, data privacy or security, or legal protection) and conceptually develop the data access regime, rests upon researchers engaging in the access process. Some of these researchers have reported various issues during the first year of data access (Jaursch et al., 2024) representing the platforms' attempts of boundary reinforcement (Griffin, 2025b). How the EC will contribute to the negotiation of acceptable data access largely

1 which means individual application forms and access modalities for each platform (Cameron Hickey et al., 2024; Casas et al., 2025; Fabio Giglietto & Massimo Terenzi, 2024)

remains to be seen, as most proceedings against, especially American, platforms' non-compliance are yet be concluded, and no larger research funding has been awarded.

Reflexive Researcher Organisation

This shows that the success of the DSA's approach to evidence-based decision-making through data access also hinges on platform cooperation, robust governance and regulatory transparency – factors researchers can only passively influence through research examining these issues. Still, the generation of knowledge to not only inform policy decisions but also the wider public, requires reflexive researcher organisation. This means, to achieve their shared interests, researchers and counter the DSA's technocratic tendencies, researchers must recognise their position in the broader political and regulatory context, and act collaboratively to effectively opening up (Stirling, 2007) discourse and practice around systemic risk. To this send, coordination (Provan & Kenis, 2007) and intermediation (Howells, 2006) promise a more efficient use of resources, increased capacity to plan for and address complex problems, higher quality of outcomes[2], and improved knowledge and information processing and brokering or the development, testing & validation of technologies and standards[3] respectively.

References

Buri, I., & van Hoboken, J. (2022, June 24). The DSA supervision and enforcement architecture—DSA Observatory. *DSA Observatory*. https://dsa-observatory.eu/2022/06/24/the-dsa-supervision-and-enforcement-architecture/

Casas, A., Dagher, G., & O'Loughlin, B. (2025). Academic access to social media data for the study of political online safety. *SocArXiv*. https://doi.org/10.31235/osf.io/7pcjd

Centeno, M. A., Nag, M., Patterson, T. S., Shaver, A., & Windawi, A. J. (2015). The emergence of global systemic risk. *Annual Review of Sociology, 41*(1), 65–85. https://doi.org/10.1146/annurev-soc-073014-112317

2 See for example the case of Democracy Reporting International against X.com resulting in important clarifications through legal precedent, or the joint analysis of VLOPSEs' first risk assessment reports by the DSA Civil Society Coordination Group, consisting of 40 different CSOs.

3 See for example EDMO's Working Group on Platform-to-Researcher Data Access, the Social Data Science Alliance initiated by the Social Media Governance Project at the Global Partnership on AI, and initiatives such as the DSA Observatory, the DSA research networ, or the DSA40 Data Access Collaboratory which serve as hubs for information exchange and generation.

De Bandt, O., & Hartmann, P. (2000). *Systemic risk: A survey* (Working Paper Series). European Central Bank.

European Digital Media Observatory. (2022). *Report of the European Digital Media Observatory's Working Group on Platform-to-Researcher Data Access* (p. 182). European Digital Media Observatory. https://edmo.eu/wp-content/uploads/2022/02/Report-of-the-European-Digital-Media-Observatorys-Working-Group-on-Platform-to-Researcher-Data-Access-2022.pdf

European Parliament and Council of the European Union. (2022). *Digital Services Act* (No. (EU) 2022/2065). European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022R2065

Fabbri, M. (2025). The role of requests for information in governing digital platforms under the Digital Services Act: The case of X. *Journalism and Media, 6*(1), 41. https://doi.org/10.3390/journalmedia6010041

Giglietto, F., & Terenzi, M. (2024). *PROMPT – The state of social media research APIs & tools in the Digital Service Act era* (No. ENO-PROMPT LC-02629302). https://drive.google.com/file/d/1htMFVxELz2hHPTCDARlWe2iqcWQsM6eQ/view

Goanta, C., Zannettou, S., Kaushal, R., van de Kerkhof, J., Bertaglia, T., Annabell, T., Gui, H., Spanakis, G., & Iamnitchi, A. (2025). The great data standoff: Researchers vs. platforms under the Digital Services Act (No. arXiv:2505.01122). *arXiv*. https://doi.org/10.48550/arXiv.2505.01122

Griffin, R. (2025a). A stakeholder mapping and research agenda: The politics of risk in the Digital Services Act. *Weizenbaum Institute*. https://doi.org/10.34669/WI.WJDS/5.2.6

Griffin, R. (2025b). Governing platforms through corporate risk management: The politics of systemic risk in the Digital Services Act. *European Law Open*, 1–31. https://doi.org/10.1017/elo.2025.17

Helbing, D. (2013). Globally networked risks and how to respond. *Nature, 497*(7447), 51–59. https://doi.org/10.1038/nature12047

Hickey, C., Dowling, K., Navia, I., & Pershan, C. (2024). *Public data access programs: A first look* (p. 51). Mozilla Foundation. https://assets.mofoprod.net/network/documents/Public_Data_Access_Programs__A_First_Look.pdf

Howells, J. (2006). Intermediation and the role of intermediaries in innovation. *Research Policy, 35*(5), 715–728. https://doi.org/10.1016/j.respol.2006.03.005

Jaursch, J., Ohme, J., & Klinger, U. (2024). Enabling research with publicly accessible platform data: Early DSA compliance issues and suggestions for improvement. *Weizenbaum Institute*. https://doi.org/10.34669/WI.WPP/9

Leerssen, P. (2024). Outside the black box: From algorithmic transparency to platform observability in the Digital Services Act. *Weizenbaum Journal of the Digital Society, 4*(2). https://doi.org/10.34669/WI.WJDS/4.2.3

Mast, T. (2024). Forschungsdatenzugang und Technologieregulierung. *Wissenschaftsrecht, 57*(2), 101. https://doi.org/10.1628/wissr-2024-0011

Provan, K. G., & Kenis, P. (2007). Modes of network governance: Structure, management, and effectiveness. *Journal of Public Administration Research and Theory, 18*(2), 229–252. https://doi.org/10.1093/jopart/mum015

Rieder, B., & Hofmann, J. (2020). Towards platform observability. *Internet Policy Review, 9*(4). https://doi.org/10.14763/2020.4.1535

Social Data Science Alliance. (2025). *The Social Data Science Alliance*. https://social-data-science-alliance.org/

Stirling, A. (2007). "Opening up" and "closing down": Power, participation, and pluralism in the social appraisal of technology. *Science, Technology, & Human Values, 33*(2), 262–294. https://doi.org/10.1177/0162243907311265

Valkenburg, P. M., Van Der Wal, A., Siebers, T., Beyens, I., Boeschoten, L., & Araujo, T. (2024). It is time to ensure research access to platform data. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-024-02066-5

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Panel 5

Session 13: Digital News

Session 14: Democratic Liberation

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Exploring Democracy-Friendly News Navigation

Konstantin Lackner \\ University of Kassel \\ Kassel \\ Germany
konstantin.lackner@uni-kassel.de

Viktoria Horn \\ University of Kassel \\ Kassel \\ Germany
viktoria.horn@uni-kassel.de

Markus Uhlmann \\ University of Kassel \\ Kassel \\ Germany
markus.uhlmann@uni-kassel.de

KEYWORDS
recommender systems; news; democracy; symbolic representation; mapping of controversies; participatory design; algorithmic experience

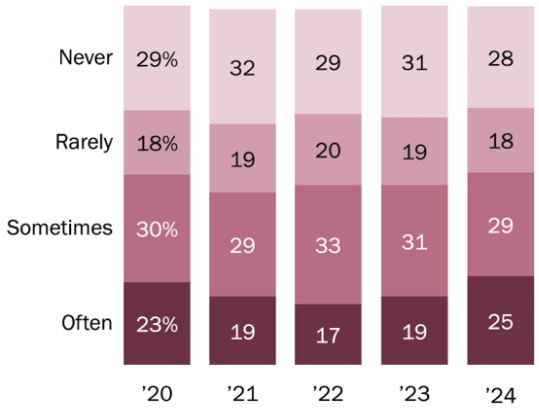
Abstract

News Recommendation is prone to be privacy-harming, opaque and misaligned with democratic values. In this paper, we elaborate on a novel concept for navigating news online beyond the logic of recommendation. Thereby we build on existing research that bridges democratic theory and user centricity for the development of an interactive visualization of articles and their respective argumentative perspectives that can be used for navigating news online. Eventually, we propose a co-design workshop with stakeholders from the field of online news alongside users in order to investigate the development of a usable prototype for news navigation. Overall, the goal is the integration into the concept of a privacy-preserving and democracy-friendly news platform for the common good in the realm of the research project Democracy-X.

Introduction

Today, a big share of news consumption happens through social media (ct. Figure 1). Thus, news consumption behaviour has shifted from consuming news in journalistically curated (online) newspapers to the logic of social media feeds (Eisenegger, 2021).

Figure 1: News consumption on social media, % of U.S. adults who get news from social media; (PEW Research Center, Social Media and News Fact Sheet 2024)



Note: Figures may not add up to 100% due to rounding.
Respondents who do not use the internet did not receive this question; they are included with those who said "Never," along with those who do not say they get news from digital devices.
Source: Survey of U.S. adults conducted July 15-Aug. 4, 2024.

Characterised by an overwhelming amount of content from different sources, users are often presented content by Recommender Systems (RS) based on metrics of popularity or their "best fit" matching e.g. previous behaviour (Mitova et al., 2023). This comes with a myriad of issues, such as a lack of a shared public sphere, potential polarisation and fragmentation of the public discourse (Flaxman et al., 2016; Helberger, 2019; Spindle et al., 2020), less diverse information exposure (Mattis et al., 2024) and user's needs such as privacy, autonomy and accuracy (Helberger, 2021, 31) to be neglected. It shows that journalism's democratic duty needs to be reconciled with practical user issues, e.g. providing more opportunities for user control (Mitova et al. 2023) and transparency "about the editorial logic behind recommendations" (Helberger, 2019, 1004).

To conclude, opinion formation in online media settings is significantly influenced by RS and thus represents a design challenge that needs to both advocate for the user's needs and the democratic mission of news media. In the research project Demokratie-X,1 an interdisciplinary research team investigates how to create a privacy-preserving, democracy-friendly online news platform. In the course of this project, the need for an alternative to the navigation paradigms introduced by social media platforms has become apparent. In particular, the research groups from the field of Participatory IT Design and Sociological Theory explore possibilities for designing democracy-friendly news navigation.

1 "Demokratie-X: Analyse der Tragfähigkeit einer privatheitsschonenden, fairen und gemeinwohlorientierten Plattform für Nachrichten" [Democracy-X: Analysis of the viability of a privacy-friendly, fair and common good-oriented platform for news], funded by the Federal Ministry of Research, Technology and Space under Grant No. 16KIS2174

Designing News Navigation Beyond the Logic of Recommendation

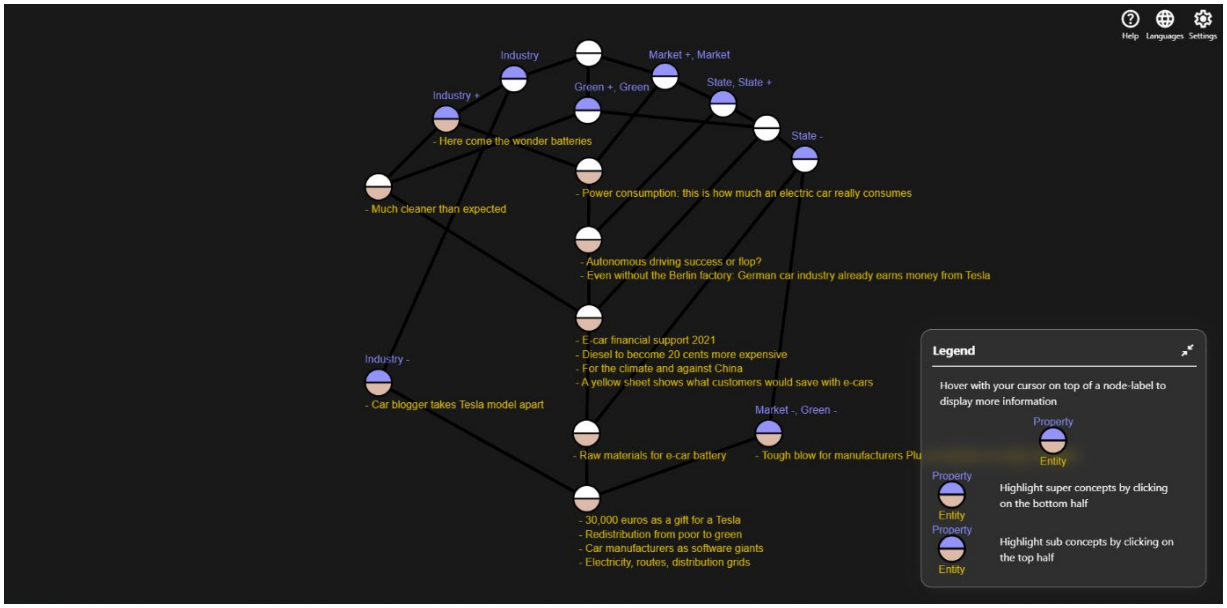
Beyond the Logic of Recommendation

The specific logic of recommendation relates to processes of selecting content on the basis of, for instance, behavioral predictions. Although recommendations can be designed against the background of democratic values (Helberger, 2019), they often do not use the specific content of journalistic articles but behavioral metrics to operationalize, for example, content diversity (Heitz et al., 2022). To counter such opaque and potential privacy harming designs, we present an approach that does not refer to the logic of recommendation in the classical sense but rather aims to enable users to navigate through public controversies (Uhlmann et al., 2025). This perspective draws on pragmatistic theories of democracy, which emphasize public disputes and justifications around complex issues (such as digitization or climate change) (Latour, 2012; Venturini & Munk, 2021). The appropriate handling of such „matters of concern" (Latour, 2012) presupposes new media formats for journalism that allow navigation through controversies and which accordingly go beyond the classical logics of news recommendation. In the following, we describe first experiences and further developments regarding the design process of a new navigation format.

First Iteration of Designing an Alternative Navigation Paradigm

To develop a prototype for news navigation, we have exemplary mapped the various forms of justification and criticism in the context of the controversy surrounding electromobility that are used in journalistic articles (Uhlmann et al., 2025). In a first iteration, we designed a prototype as web-application that visualizes various argumentative perspectives as well as their specific connections in journalistic articles, while also allowing an easy identification of the least and most diverse articles (see Figure 2). Results of a first user study show that this form of navigation is attractive in terms of stimulation for discovering the diversity of perspectives (Horn et al., 2024). However, we also identified research challenges such as trade-offs between visualizing the complexity of controversies and usability as well as challenges regarding proficient business models that use the novel navigation form.

Figure 2: Screenshot of the prototype of the web-application that visualizes various argumentative perspectives connected to their respective news articles; taken from Horn et al., 2024



Co-Design Workshop

Based on the introduced concept, we plan a participatory study for prototyping a usable democracy-friendly news navigation that addresses the previously mentioned design challenges. For this, we will facilitate a Co-design Workshop as outlined by (Steen, 2013). Through Demokratie-X, we have already established contact with different stakeholders from the field of online news media, such as journalists and their respective publishers, which we want to include in the participatory design process alongside prospective users of such a news platform.

Discussion and Outlook

Our contribution empowers people to critically assess their online news selection practices by offering a new approach for news navigation. Additionally, by including multiple stakeholder perspectives in the design process, we take into account the complex ecosystem of online news media and its interplay of different logics and values.

References

Eisenegger, M. (2021). Dritter, digitaler Strukturwandel der Öffentlichkeit als Folge der Plattformisierung. In M. Eisenegger, M. Prinzing, P. Ettinger, & R. Blum (Eds.), *Digitaler*

Strukturwandel der Öffentlichkeit: Historische Verortung, Modelle und Konsequenzen (pp. 17–39). Springer Fachmedien. https://doi.org/10.1007/978-3-658-32133-8_2

Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly, 80*(S1), 298–320.

Heitz, L., Lischka, J. A., Birrer, A., Paudel, B., Tolmeijer, S., Laugwitz, L., & Bernstein, A. (2022). Benefits of diverse news recommendations for democracy: A user study. *Digital Journalism, 10*(10), 1710–1730. https://doi.org/10.1080/21670811.2021.2021804

Helberger, N. (2019). On the democratic role of news recommenders. *Digital Journalism, 7*(8), 993–1012. https://doi.org/10.1080/21670811.2019.1623700

Horn, V., Hirth, J., Holfeld, J., Behmenburg, J. H., Draude, C., & Stumme, G. (2024). Disclosing diverse perspectives of news articles for navigating between online journalism content. In *Proceedings of the 13th Nordic Conference on Human-Computer Interaction* (pp. 1–14). https://doi.org/10.1145/3679318.3685414

Latour, B. (2012). *Das Parlament der Dinge*. Suhrkamp.

Mattis, N., Masur, P., Möller, J., & van Atteveldt, W. (2024). Nudging towards news diversity: A theoretical framework for facilitating diverse news consumption through recommender design. *New Media & Society, 26*(7), 3681–3706. https://doi.org/10.1177/14614448221104413

Mitova, E., Blassnig, S., Strikovic, E., Urman, A., Hannak, A., de Vreese, C. H., & Esser, F. (2023). News recommender systems: A programmatic research review. *Annals of the International Communication Association, 47*(1), 84–113. https://doi.org/10.1080/23808985.2022.2142149

Pew Research Center. (2024, September 17). *Social media and news fact sheet.* https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/

Spinde, T., Hamborg, F., Donnay, K., Becerra, A., & Gipp, B. (2020). Enabling news consumers to view and understand biased news coverage: A study on the perception and visualization of media bias. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020* (pp. 389–392). Association for Computing Machinery. https://doi.org/10.1145/3383583.3398619

Steen, M. (2013). Co-design as a process of joint inquiry and imagination. *Design Issues, 29*(2), 16–28. https://doi.org/10.1162/DESI_a_00207

Uhlmann, M., Hirth, J., & Horn, V. (2025). Jenseits der Logik der Empfehlung: Formale Begriffsanalyse als Grundlage für eine neue Variante zur Vermittlung von Nutzenden und journalistischen Inhalten. In K. Krenn, J. Kropf, S. Laser, & C. Ochs (Eds.), *Dynamiken digitaler Bewertung. Über Gestaltungsspielräume in Infrastrukturen – von KI bis Queering* (pp. 137–155). Springer. https://doi.org/10.1007/978-3-658-46989-4_7

Venturini, T., & Munk, A. K. (2021). *Controversy mapping: A field guide.* Polit

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Mapping Germany's #Remigration Debate through the Lens of Social Media

A Comparative Study of Facebook and Instagram

Julian Maitra \\ University of Fribourg \\ Fribourg \\ Switzerland
julimaitra@gmail.com

Anas Ansar \\ Arnold Bergstraesser Institut, Universität Freiburg \\ Freiburg \\ Germany
anas.ansar@abi.uni-freiburg.de

KEYWORDS
remigration; Facebook; Instagram; political communication; platformization; polarization, social media

Introduction

This study investigates the far-right-driven "Remigration" debate in Germany, which intensified after Correctiv's January 2024 investigation (Bornmann, 2024) on AfD politicians' involvement in extremist discussions about mass deportation. Termed *Unwort* (worst word) of 2023, "Remigration" has polarized public discourse, particularly on Meta-owned platforms Facebook and Instagram. Combining mixed methods, we analyze 9,896 posts (8,056 Facebook, 1,840 Instagram) collected via CrowdTangle from February 2023–February 2024, totaling 3.3 million user interactions. The research explores temporal engagement patterns, content dynamics, and societal implications of this debate.

Background

In German, Remigration traditionally denotes voluntary return migration, as outlined in the Handbook of Return Migration (King & Kuschminder, 2022). However, far-right actors have co-opted the term to frame forced deportation of migrants, asylum seekers, and naturalized citizens, intertwining it with ethnonationalist and racist rhetoric (Akerlund, 2021; Laaksonen et al., 2020). This debate reflects broader anxieties about identity and integration, amplified by social media's role in

normalizing extremist narratives (Papacharissi, 2015). Ahmed's (2004) concept of "cumulative racism" explains how fringe ideas gain mainstream traction through repeated in-group/out-group dynamics, a process accelerated by platforms' algorithmic architectures (Matamoros-Fernandez, 2017; Fekete, 2014).

Methodology

The study employs a mixed-methods approach. Quantitatively, we apply exploratory analysis of social media trace data (Freelon, 2014) to map temporal trends in post frequencies and user interactions. Text-mining techniques and hashtag co-occurrence networks reveal content patterns. Qualitatively, interpretative content analysis focuses on high-engagement posts to unpack framing strategies. This dual approach bridges computational methods with critical discourse analysis, addressing platforms' role in shaping migration narratives through "civil" and "uncivil" discursive practices (Krzyzanowski & Ledin, 2017; Ekman, 2019). The study addresses the following RQs:

RQ1: How do temporal trends in post volume and user engagement (e.g., likes, comments) around "Remigration" differ between Facebook and Instagram?

RQ2: What content types (e.g., partisan narratives, news articles) drive high engagement on each platform?

RQ3: How does the social media debate on Remigration reflect or influence broader societal polarization and democratic discourse in Germany?

Key Findings

Post volumes surged during the January 2024 AfD scandal, with Instagram interactions (likes/comments) rising sharply, while Facebook engagement remained stable despite similar posting rates. This aligns with Yarchi et al.'s (2021) observation of "affective polarization" during crises.

Right-wing actors, particularly AfD accounts, dominated posting frequencies on both platforms. However, Instagram's younger, visually oriented user base amplified counter-narratives from left-wing and news accounts, achieving higher engagement with fewer posts. Hashtag networks revealed polarized clusters (pro/anti-Remigration), weakly linked via #afd, reflecting Stier et al.'s (2017) findings on AfD's social media influence.

The debate exemplifies how algorithmic amplification fosters "cumulative racism" (Ahmed, 2004), legitimizing extremist rhetoric. Facebook's shareability reinforces echo chambers (Heidenreich et al., 2020), whereas Instagram's affordances enable marginalized voices to challenge dominant narratives (Stravato Emes, 2024). These dynamics mirror global trends where anti-immigration sentiments bolster right-wing populism (Van Heerden et al., 2014; Walgrave & De Swert, 2004).

This study aligns with the conference's focus on democracy and well-being in digital societies by:

- Examining platform architectures: Highlighting how Meta's algorithms amplify anti-democratic narratives, undermining social cohesion.
- Interplay of agency and curation: Demonstrating how user interactions and algorithmic filtering co-shape polarized discourse.
- Policy implications: Advocating for regulatory frameworks that balance free speech with protections against hate speech, informed by Krzyżanowski and Ledin's (2017) work on discursive governance.

Conclusion

The Remigration debate underscores social media's dual role as a catalyst for democratic engagement and a vector for illiberal ideologies. Policymakers must address platform governance to mitigate harms while preserving civic participation. Future research should explore intersectional vulnerabilities, particularly for marginalized groups, to foster inclusive digital spaces.

References

Åkerlund, M. (2021). Influence without metrics: Analyzing the impact of far-right users in an online discussion forum. *Social Media + Society, 7*(2), 20563051211008831. https://doi.org/10.1177/20563051211008831

Bornmann, M. (2024). Geheimplan gegen Deutschland. *Correctiv.org*. https://correctiv.org/aktuelles/neue-rechte/2024/01/10/geheimplan-remigration-vertreibungafd-rechtsextreme-november-treffen

Diehl, C., & Liebau, E. (2015). Turning back to Turkey–or turning the back on Germany? Remigration intentions and behavior of Turkish immigrants in Germany between 1984 and 2011. *Zeitschrift für Soziologie, 44*(1), 22–41.

Ekman, M. (2019). Anti-immigration and racist discourse in social media. *European Journal of Communication, 34*(6), 606–618.

Fekete, L. (2014). Europe against the Roma. *Race & Class, 55*(3), 60–70.

Freelon, D. (2014). On the interpretation of digital trace data in communication and social computing research. *Journal of Broadcasting & Electronic Media, 58*(1), 59–75.

Heidenreich, T., Eberl, J. M., Lind, F., & Boomgaarden, H. (2020). Political migration discourses on social media: A comparative perspective on visibility and sentiment across political Facebook accounts in Europe. *Journal of Ethnic and Migration Studies, 46*(7), 1261–1280.

King, R., & Kuschminder, K. (Eds.). (2022). *Handbook of return migration*. Edward Elgar Publishing.

Krzyżanowski, M., & Ledin, P. (2017). Uncivility on the web: Populism in/and the borderline discourses of exclusion. *Journal of Language and Politics, 16*(4), 566–581.

Laaksonen, S. M., Pantti, M., & Titley, G. (2020). Broadcasting the movement and branding political microcelebrities: Finnish anti-immigration video practices on YouTube. *Journal of Communication, 70*(2), 171–194.

Matamoros-Fernández, A. (2017). Platformed racism: The mediation and circulation of an Australian race-based controversy on Twitter, Facebook and YouTube. *Information, Communication & Society, 20*(6), 930–946.

Nordbrandt, M. (2023). Affective polarization in the digital age: Testing the direction of the relationship between social media and users' feelings for out-group parties. *New Media & Society, 25*(12), 3392–3411.

Papacharissi, Z. (2015). Affective publics: Sentiment, technology, and politics. Oxford University Press.

Stier, S., Maier, J., & Oschatz, C. (2021). Die Nutzung sozialer Medien durch Kandidaten im Bundestagswahlkampf 2017. In *Wahlen und Wähler: Analysen aus Anlass der Bundestagswahl 2017* (pp. 415–432).

Stravato Emes, C. (2024). Is it about "them"? Leveraging big data research to understand anti-immigrant discourse. *Big Data & Society, 11*(2), 20539517241249432. https://doi.org/10.1177/20539517241249432

Van Heerden, S., De Lange, S. L., van der Brug, W., & Fennema, M. (2014). The immigration and integration debate in the Netherlands: Discursive and programmatic reactions to the rise of anti-immigration parties. *Journal of Ethnic and Migration Studies, 40*(1), 119–136.

Walgrave, S., & De Swert, K. (2004). The making of the (issues of the) Vlaams Blok. *Political Communication, 21*(4), 479–500.

Yang, G. (2016). Narrative agency in hashtag activism: The case of #BlackLivesMatter. *Media and Communication, 4*(4), 13–17.

Yarchi, M., Baden, C., & Kligler-Vilenchik, N. (2021). Political polarization on the digital sphere: A cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. *Political Communication, 38*(1–2), 98–139.

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Shifting Discursive Alliances

A Longitudinal Analysis of Australian Climate Change Discourses on Facebook through Practice Mapping

Axel Bruns \\ Queensland University of Technology \\ Brisbane \\ Australia
a.bruns@qut.edu.au

Carly Lubicz-Zaorski \\ Queensland University of Technology \\ Brisbane \\ Australia
carly.lubicz@hdr.qut.edu.au

Tariq Choucair \\ Queensland University of Technology \\ Brisbane \\ Australia
tariq.choucair@qut.edu.au

Laura Vodden \\ Queensland University of Technology \\ Brisbane \\ Australia
laura.vodden@qut.edu.au

Ehsan Dehghan \\ Queensland University of Technology \\ Brisbane \\ Australia
e.dehghan@qut.edu.au

KEYWORDS
climate change; discursive alliances; Facebook; Australia; practice mapping

Introduction

As global heating continues, public debates about climate change also shift. Faced with mounting evidence of more frequent and extreme weather events around the globe, narratives opposing climate action have moved from outright climate change denial to delay tactics (Lamb et al., 2020; Painter et al., 2023). Conversely, advocates for climate action are making increasingly forceful arguments for urgent practical interventions.

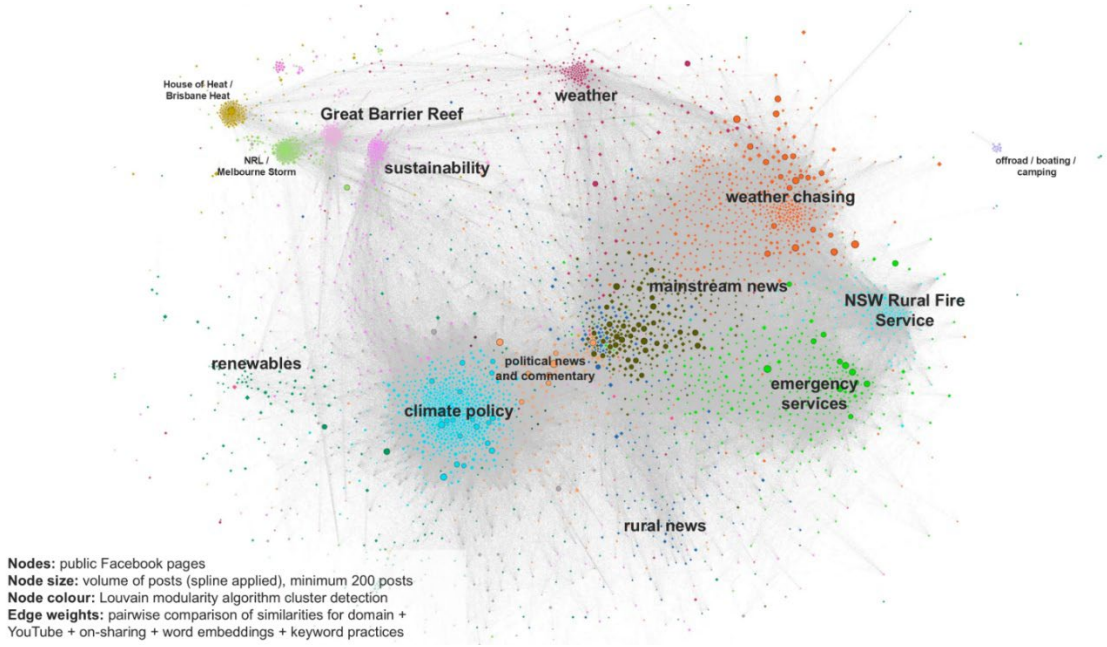
Such discursive strategies can be observed in many nations, yet Australia makes for an especially useful case study: it is particularly exposed to the consequences of climate change, has experienced increasingly severe disasters including cyclones, floods, and bushfires across its diverse climatic zones; and features a parliamentary political system which spans the full breadth from outright climate change denial (amongst the conservative Coalition) through support for modest incremental

initiatives (from Labor) to calls for urgent action (by the Greens). Additionally, a new group of 'Teal' independents (named after their campaigning colours) blending pro-business and environmental values has successfully challenged incumbent MPs perceived to be stalling climate action (Hendriks & Reid, 2024).

Data and Methods

This paper examines the evolution of Australian public debate on climate action over nearly seven years, from January 2018 to August 2024 (spanning significant extreme weather events in 2019/20 and a change of federal government in May 2022). We draw on data from CrowdTangle on public discussions on Facebook, filtered for public pages from Australia and for posts that contain one or more climate-related keywords. We chose Facebook because it remains the most popular social media platform in Australia (e.g. Park et al., 2022: 85); our data capture post content and engagement data from official political pages (parties, politicians, candidates), news outlets, civil society groups, activists, local community pages, and various other actors. The total dataset contains some 4.8 million posts.

Figure 1: A map of distinct practices in the dataset, highlighting groups of pages engaging in climate policy debate, broader political commentary and news reporting, emergency responses, weather chasing, and other themes. Some clusters have a specific local or regional focus.



We conduct a longitudinal analysis of this dataset using the practice mapping method (Bruns et al., 2025), which is especially useful in this context as CrowdTangle data contain no interaction network information. Practice mapping instead constructs networks between individual Facebook pages by systematically comparing them for similarities in their posting practices: this includes general language choices, specific climate change claims, references to other actors and entities, embedded links, images, and videos, and other discursive features that can be extracted from the post content

– collectively, their posting practices. Pages are then clustered into larger groups based on the strength of affinities between their practices, which in turn also enables us to determine the relative alignment or opposition between these larger clusters. Fig. 1 shows a preliminary practice mapping network for our dataset, using only a limited set of content features, illustrating the detection of distinct discursive clusters.

Practice mapping networks typically show several distinct clusters, featuring closely aligned actors that represent a particular discursive position; these clusters are in turn often grouped into larger discursive alliances that each stand for a broader agenda (e.g. for or against immediate climate action), and engage in an antagonistic and potentially polarised discursive struggle with each other.

Contribution

We construct such a practice mapping network for the entire multi-year dataset, identifying broad overall patterns of agonism and antagonism (Dehghan, 2020) in Australian climate debates, but also trace the positioning of individual pages in our dataset across this map over time – month by month and year by year. This charts the diachronic evolution of climate change discourses in Australia: actors originally engaged in outright climate change denialism might move towards climate action delay as denial becomes untenable in the face of the evidence; actors calling for modest global action might move towards support for more urgent local initiatives.

We pay particular attention to the impact of major unforeseen developments (bushfire and flood emergencies; protest actions) and regular events (annual climate summits; release of scientific reports; federal elections in 2019 and 2022) on discursive practices. We cross-reference these practices with user engagement metrics for pages and posts, to examine whether audience preferences for specific stances towards climate change also evolve over time, and potentially even whether this evolution precedes or lags behind discursive changes at the page level.

Applying practice mapping to a large-scale longitudinal dataset, this paper makes a unique contribution to the study of discursive shifts in public debate at a national level. Our work documents the diachronic contingency of discourses on endogenous or exogenous events; polarisation dynamics around climate action in Australia; and the resonance of arguments for or against urgent climate action amongst the broader public. It also serves as a blueprint for similar longitudinal studies of public debates on and across other platforms.

References

Bruns, A., Kasianenko, K., Suresh, V. P., Dehghan, E., & Vodden, L. (2025). Untangling the Furball: A Practice Mapping Approach to the Analysis of Multimodal Interactions in Social Networks. *Social Media + Society*, *11*(2). https://doi.org/10.1177/20563051251331748

Dehghan, Ehsan. 2020. *Networked Discursive Alliances: Antagonism, Agonism, and the Dynamics of Discursive Struggles in the Australian Twittersphere*. PhD thesis. Queensland University of Technology. https://doi.org/10.5204/thesis.eprints.174604

Hendriks, Carolyn M., and Richard Reid. 2024. "Citizen-Led Democratic Change: How Australia's Community Independents Movement Is Reshaping Representative Democracy." *Political Studies* 72 (4): 1609–1631. https://doi.org/10.1177/00323217231219393

Holmes à Court, Simon. 2023. *The Big Teal*. Melbourne: Monash University Publishing.

Lamb, William F., Giulio Mattioli, Sebastian Levi, J. Timmons Roberts, Stuart Capstick, Felix Creutzig, Jan C. Minx, Finn Müller-Hansen, Trevor Culhane, and Julia K. Steinberger. 2020. "Discourses of Climate Delay." *Global Sustainability* 3: e17. https://doi.org/10.1017/sus.2020.13

Painter, James, Joshua Ettinger, David Holmes, Loredana Loy, Janaina Pinto, Lucy Richardson, Laura Thomas-Walters, Kjell Vowles, and Rachel Wetts. 2023. "Climate Delay Discourses Present in Global Mainstream Television Coverage of the IPCC's 2021 Report." *Communications Earth & Environment* 4 (1): 1–12. https://doi.org/10.1038/s43247-023-00760-2

Park, Sora, Kieran McGuinness, Caroline Fisher, Jee Young Lee, Kerry McCallum, and David Nolan. 2022. *Digital News Report: Australia 2022*. Canberra: News and Media Research Centre. https://apo.org.au/node/317946

Empowering People in Online Spaces:
Democracy and Well-Being in Digital Societies

Democracy by Default

Hatching a Commons for Digital Deliberation

Torsten Goerke \\ TU Dresden | CODIP \\ Dresden \\ Germany
torsten.goerke@tu-dresden.de

Sabine Barthold \\ TU Dresden | CODIP \\ Dresden \\ Germany
sabine.barthold@tu-dresden.de

KEYWORDS
digital democracy; ontologies; protocols; deliberation; governance; digital commons

Introduction

In this paper, we present a framework that uses a protocol-based approach to transform ontologies of social norms into actionable agendas for the development of democratic technical systems. We discuss how the democratic governance of communities, community-owned infrastructures, and digital commons can be organized based on dynamic and fluid deliberation processes using socio-technical protocols. We examine how communities can be empowered to control algorithms, data, and infrastructure through systems that are "democratic by design," focusing on the emerging vision of digital democracy where collaborative technologies form a symbiotic relationship with democratic processes. After introducing key principles for democratic technologies and governance processes, we propose the concept of hatching, a process that is sensitive to the cultural variances, different technological visions, and dynamic technology needs for intersecting social groups. This concept of a commons for digital deliberation helps reimagine the development of technologies and infrastructures as a community-oriented instead of user-oriented approach based on flexible protocols instead of stack models.

Technological Challenges to Democracy

In today's digital society the technical systems that enable participation in public discourse, access to knowledge and information, and building communities are increasingly relying on proprietary infrastructure governed by opaque entanglements between humans and algorithms. This

entanglement has increased with the development of large language models (LLMs) that shape user experiences as a "multiverse of echo chambers" (Lu, 2023), giving rise to the "algorithmic internet." People have less power over the design and governance of these infrastructures and their own data and algorithms. Two major interrelated trends signify this anti-democratic shift in technology development: (1) *Anti-social trends,* driven by technologies like social media, crypto or FinTech, break down the social fabric, erode norms, heighten polarization and lead to an expanding reach of finance technologies within societies and are largely unaccountable to democratic principles and control. The second thread is (2) the *centralization trend*: new technologies (machine learning, foundation models, IoT) tend to centralize power in the hands of actors with privileged access to both data and capital so that "small groups of engineers can have the ability to set patterns in systems that shape the rules of social life for billions of citizens" (Weyl & Tang, 2024) These centralizing trends increasingly normalize unprecedented systems of surveillance and centralized control over information (Zuboff, 2019). These technological trends are putting increasing pressure on democratic societies when being used by anti-democratic movements and they strengthen authoritarian regimes as means of power and control. To overcome these trends, technology must be understood as politics, with different visions and options available for a technological future.

Digital Democracy: Key Principles

To overcome these anti-democratic trends, we suggest three key principles that digital democracy needs to address: Firstly, in a digital democracy, technologies help empower emerging publics. According to John Dewey (1927), all new technologies are corresponding to *emerging publics* with their own spatial or temporal interdependencies, which create challenges to democratic processes. Digital technologies and the *emerging publics* they help create today are mostly managed by markets and often poorly represented or governed by preexisting (democratic) institutions. A democratic governance approach to digital technologies would require that new challenges posed by these socio-technical systems should be managed by the "relevant public", meaning groups of people whose lives are directly affected and shaped by this technology. Secondly, digital democracy accounts for the plurality of intersecting social groups. People's identities are constituted by intersecting communities that form the core fabric of the social world (Simmel 1908). So, democratic technologies enable networked and intersecting identities and encourage dynamic group relations, societal complexity and difference, a vision for which Weyl and Tang (2024) have coined the term *Plurality*. Plurality empowers individuals to form and express different group identities and participate in intersecting communities with their distinct social relations, identity systems, decision-making procedures, and property regimes. And thirdly, as a direct result of the erosive, anti-democratic, and surveillance-increasing tendencies, many people have developed a general skepticism and discontent with technological progress. Digital democracy needs to formulate a positive vision of progress in which collaborative technologies are forming a symbiotic relationship with democratic practices. Reimagining digital technologies as public goods and services and increasing public investment in their infrastructures can be a first step towards digital democracy. But it needs a paradigm shift to make democratic principles the default in technology development.

Democratic Technology: A New Ontology

An ontology is a formal representation of a set of concepts and their relationships within a given domain. The challenge in building democratic socio-technical systems is often rooted in the translation of different ontologies and vocabularies between the engineering, social sciences, and humanities (Callon, 1987). Technical systems rely on ontologies of language, objects, and functions. Social systems rely on concepts of power, communication, identity, and relationships, as well as culture and history. Understanding these two realms of knowledge, not as separate systems with their own ontology, but instead as assemblages, helps to shift the focus on associations and relations between human and nonhuman actants in the study of technology. Ontologies of social norms can then provide a structured and organized framework for understanding the complex web of social guidelines that govern human behavior in association with technology.

Hatching the Commons

In this paper we propose that instead of thinking of technical systems in terms of stacks that are built by software architects based on presupposed user needs, we should establish processes that help develop technical systems that are rooted in dynamic communities, their cultural practices, and intersecting identities, as well as their socio-technological visions. Protocols can help transform stack-oriented thinking into a more organic and dynamic way to build systems that serve communities' needs. Deschermayer (2022) suggested the term "Hatch" to indicate this sensitivity to the initial social conditions that precede any building process. Hatching as a cultural build is grounded in principles for commons stewardship (Ostrom, 1990) and the nurturing of cultural processes. Hatching is the initial phase in creating the Commons: "Engineering any complex system is a consideration of its initialization conditions—in a Commons, we call these initial conditions the 'Hatch'. […] because new paradigms sometimes require new terminology to avoid misconceptions and the connotations of old vocabulary" (Commons Stack, 2021).

Hatching introduces the new role of a governor-engineer (Zhang, 2025), which is crucial for transforming a theoretical social framework into actionable steps and cultivating a process of techno-scientific governance (Fritsch, 2020). A governor-engineer is a person who can apply knowledge from building technical systems in the realm of governance and helps shift from the practice of requirements engineering, where requirements are translated into features in the process of software development, to a new process that creates actual agency among communities and individual users (or inhabitants of digital spaces) by defining them as pioneers, leaders, facilitators, or founders, etc.: those who educate, those who explain new interdependencies to others, and those who help these "new publics" to come into existence are engaging in systems building as a deliberative practice.

Protocols for Democratic Socio-Technical Systems – ATProto as an Illustrating Example

Protocols can function as a method to translate between different realms of thinking and to find solutions to coordination problems in digital society (Rao et al., 2023). The use of protocols, even though they are "hard to study and easy to ignore" (Rao et al., 2023), is in fact not such a new idea but has been part of the utopia of the internet from its very beginning: a system of frictionless global communication and instant access to the knowledge of humanity—a "dream of scale and ease" (Lu, 2023).

The Authenticated Transfer Protocol (ATProto) is a decentralized protocol for large-scale social web applications. ATproto is the protocol underlying the Bluesky social network and the IndieSky movement, which set out to create infrastructure for open social media in the jurisdiction of Europe. Central initial hatching conditions are based on two technical pillars: sovereign identity and algorithmic choice. Let's briefly look at how identity is managed on Atproto: Users in ATProto have permanent decentralized identifiers (DIDs) for their accounts, which are immutable. This means once created, they cannot be altered; all content is signed by this identifier, and the protocol is agnostic to where the data is stored. DIDs are the long-term persistent identifiers for accounts in ATproto, but they can be opaque and unfriendly for human use. The mechanism for verifying the link between an account handle and an account DID relies on DNS—the domain name system. Thus, every handle must be a valid network hostname. Handles are a less permanent but more human-friendly identifiers for accounts.

User data is exchanged in signed data repositories. These repositories are collections of records, which include posts, comments, likes, follows, etc. stored at a hosting provider. Using a configurable domain name, which acts as a human-readable handle, identities include a reference to the user's current hosting provider and cryptographic keys used to sign their data. This interplay of data servers, immutable DIDs and mutable account handles ensures that users can migrate their accounts to any new data server, and that they can do this without the server's involvement—which means they do not need to ask for permission. All data signed by their cryptographic key is tied to their ID and can move freely between data providers, and all relationships persist between such migrations. Lock-in effects are effectively removed. On BlueSky the social graph is open and managed as a public good. Everybody is invited to build applications tailored to their communities' needs on top of it. The second pillar important for the hatch (the initialization conditions) is algorithmic choice. Users on ATproto are free to select their feeds and can even choose to build their own custom feeds that others can subscribe to. Custom feed builders, like Graze, can help build custom feeds on Bluesky and allow users to build, deploy, and manage their feed, including custom timelines and sort order of posts.

Conclusion

In this paper we have discussed how digital democracy aims to transform social norms into actionable agendas for democratic technical systems by focusing on the democratic governance of communities, community-owned infrastructures, and digital commons through dynamic deliberation processes using socio-technical protocols. The concept of hatching the commons suggests a shift from traditional stack-oriented thinking to a more organic and dynamic approach, grounded in principles for commons stewardship and the nurturing of cultural processes. Protocols like ATProto can help create a socio-technological ecosystem that translates between different realms of thinking and helps to find solutions to coordination problems in digital society. When regarded in this way, protocols can help advance socio-technical systems that are inherently democratic.

However, the social and technical complexity of a transition from proprietary platforms and privately owned infrastructures remains substantial, and social lock-in effects are still powerful and unlikely to change rapidly, so that many open questions remain. For example, what are incentive structures for participation in deliberative and democratic infrastructure commons, and how can the transformation be facilitated? To address these open questions, we envision the collaborative development and maintenance of software-mediated systems as a deliberative practice grounded in interdisciplinary co-creation between intersecting communities as well as engineering and social science disciplines.

References

Callon, M. (1987). Society in the making: The study of technology as a tool for sociological analysis. In W. E. Bijker, T. Parke Hughes, & T. Pinch (Eds.), *The social construction of technological systems: New directions in the sociology and history of technology* (pp. 83–103). MIT Press.

Commons Stack. (2021, April 29). What's in a hatch? Understanding the initial phase of the Token Engineering Commons. *Medium*. https://medium.com/commonsstack/whats-in-a-hatch-3799a398243a

Deschermayer, L. (2022, February 17). Cultural build: A framework for self-actualizing economies [Conference presentation]. ETH Denver.

Dewey, J. (1927). *The public and its problems.* Holt Publishers.

Fritsch, F. (2020). The Commons Stack: Realigning incentives towards public goods [Case study]. http://dx.doi.org/10.13140/RG.2.2.12931.73767

Lu, C. (2023). The algorithmic internet: Culture, capture, corruption. In *Paradigm Trilogy II: Man vs Machine 2*. https://www.paradigmtrilogy.com/assets/documents/issue-02/christina-lu--the-algorithmic-internet.pdf

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.

Rao, V., Beiko, T., Ryan, D., Stark, J., Van Epps, T., & Aue, B. (2023, March 9). The unreasonable sufficiency of protocols. *Summer of Protocols (blog)*. https://venkatesh-rao.gitbook.io/summer-of-protocols

Simmel, G. (1908). *Soziologie: Untersuchungen über die Formen der Vergesellschaftung*. Duncker & Humblot.

Swan, A.-M. (2021, August 24). Commons: The heart of the creation era. *Commons Stack (blog)*. https://medium.com/commonsstack/commons-the-heart-of-the-creation-era-c4b1ccdb864d

Weyl, G., Tang, A., & ⿻ Community. (2024, May 2). *Plurality: The future of collaborative technology and democracy*. https://www.plurality.net/chapters/

Zhang, G. Z. (2025, January 6). The critical legacy of Chinese cybernetics. *Combinations (blog)*. https://www.combinationsmag.com/the-critical-legacy-of-chinese-cybernetics/

Zuboff, S. (2019*). The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Public Affairs