

TWON Policy Brief #2

On Regulating Online Social Networks

January 2025

Executive Summary

In the digital age of individualization and digitalization, communities have grown further apart and digital communication has partly replaced connections within one's physical neighborhood. The promise of wider networks facilitated by social media is often at odds with the aim of generating in-depth, meaningful deliberation where people sincerely listen and learn from each other. Democracies face significant challenges, such as the spread of disinformation, hate speech, and polarization on digital platforms. As in the case of the 2024 presidential election in Romania, foreign actors such as the Russian government, are furthermore suspected of influencing elections by supporting a candidate's campaign on TikTok. It has been warned that these issues erode public trust, undermine the inclusiveness of democratic decision making and the learning potential of our societies. Despite these threats, digital platforms also present opportunities for increased democratic participation. This policy brief outlines strategies to build a resilient digital democracy that can mitigate the risks posed by digital platforms driven exclusively by economic motives, while enhancing opportunities for engagement. Key areas include the regulation of platforms, establishment of public/independent participatory platforms, and improvement of media literacy.

Underregulated Digital Platforms are Endangering Democracy

The Arab Spring demonstrated the democratic power of online social networks. However, changing digital media environments have created dependence on a few dominant platforms (Luca and Bazerman 2020) and their engagement-driven business models (Srnicek 2016). For example, Twitter, now X, was once known for content moderation but has seen a rise in disinformation and hate speech, such as antisemitism, following its takeover by Elon Musk (Miller et al. 2023). Despite boycotts by advertisers over the placement of ads near harmful content, X responded with a lawsuit rather than with content moderation.

While the empirical evidence is mixed, it has been warned that social media news feeds, driven by recommendation systems, contribute to polarization by reinforcing ideological bubbles and amplifying negative emotions, which lead to affective polarization. While platforms are aware of these effects, they resist changes to their algorithms, as this might endanger their business model (Ludwig et al. 2023; Bojic, 2024). The dependency on platforms, the spread of hate speech, and algorithmic bias underscore the need for political action.

The European Commission (2018) recognized the urgency of combating disinformation, especially as AI tools further amplify the problem. How can we address disinformation and polarization driven by platform providers?

Policy Options

1. **Self-Regulation by Platforms:** Self-regulation has had limited success, and challenges remain in enforcing content moderation and addressing privacy concerns related to bot and user verification.
2. **Regulation of Platform Algorithms:** Platforms must make their algorithms transparent and subject to external audits to ensure accountability and prevent harmful feedback loops. To address the challenges of content diversity and balance in online algorithmic recommendations, algorithms should be intentionally designed to deliver a curated mix of content.
3. **Public/Independent Participatory Platforms:** Decentralized platforms like Mastodon could provide more inclusive discourse, relying on pro-democratic and transparent algorithms. Publicly funded non-state entities, such as the Wikimedia Foundation, could manage these platforms to avoid state misuse.
4. **Improving Media and Data Literacy:** With AI-generated disinformation rising, targeted campaigns to improve media literacy are essential. Explainable AI tools can help users assess content validity in real time.
5. **Support for Independent Media:** EU-level funding mechanisms could support independent journalism, which is key to countering misinformation and fostering public discourse.

Policy Recommendations

1. Fund Research on OSN

To understand the complex mechanisms and effects of OSN, researcher access to platform data (such as under the DSA) is necessary, but also sufficient funds to conduct the research. We need research quantifying undesired effects like opinion polarization, affective polarization, falsehood dissemination and the impact of foreign powers. Additionally, we need to investigate how design decisions of OSN can lead to undesired effects on citizens and societies.

To enable this research, the researcher's access to platform data under the DSA needs to be fully implemented and simplified, as stated in more detail in section 3. 1.

2. Fund the Development of Public Platforms

Allocate funding to support the development of independent digital platforms that are aligned with EU standards and promote inclusive participation.

Alternatively, create mandatory public safe spaces within existing platforms, where only authenticated and high-quality regional content is displayed. These spaces could be regulated and maintained by public, non-profit actors, while at the same time using the existing infrastructure, where users profit from network effects.

Moreover, e-participation tools should be used to enable meaningful participation of citizens in local debates and actual decision-making at different levels of the state.

3. Enforce Interoperability Between Platforms

Similar to industry standards, force big platforms to ensure interoperability. This way, people are free to engage with their platform of choice, not forced towards the one with the biggest user base. This is an important base for building successful alternatives to existing platforms.

4. Increase Transparency and Accountability

Platforms must be mandated to publish detailed reports on their content moderation practices and provide external access to their algorithms to ensure they are not promoting disinformation and emotiona-lized content. Art. 40 DSA is a good basis, but researcher access needs to be simplified in practice (see section 3. 1.).

5. AI-Driven Content Verification

Invest in the development of technologies such as explainable AI to classify and debunk disinformation in real time, ensuring that users are informed about the credibility of the content they engage with.

6. Promote Media Literacy

Launch media literacy campaigns targeting various age groups, with a particular focus on empowering individuals to recognize and counteract disinformation and manipulation by AI tools.

7. Support Independent Journalism

Create a European fund to support independent media outlets that adhere to high-quality standards. This is important to foster fact-based information even on OSN and to alleviate the economic pressure on media outlets caused by the shift of advertisement budgets away from press and towards OSN. This fund should be managed by an independent body to ensure transparency and accountability.

8. Promote Content Diversity through Algorithm Design

Algorithms should be designed to deliver a curated mix of content that balances emotional tones, to avoid negative bias, and introduces users to various topic areas and political viewpoints, fostering a more inclusive and creative digital environment. To enhance personalization while preventing echo chambers, users should be empowered with customizable settings that allow them to adjust their content diversity preferences. This promotes personalized digital autonomy while ensuring a baseline exposure to differing perspectives. Tech companies, as key players in this model, should adhere to agreed standards in algorithm design and transparently demonstrate their contributions to collective decision-making processes (Bojic, 2024).

Appendix: The Process of Developing the First Policy Brief

TWONs offer crucial possibilities to measure and simulate the effects of online social networks. However, even the basic communication of the concept of a digital twin of an online social network, or, in short, a TWON, presents challenges, let alone translating interdisciplinary findings into actionable insights for decision-makers in politics and industry. To make an impact with our research, we want to face this challenge and develop policy recommendations based on scientific results. Additionally, we want to foster digital citizenship and the public debate on the role Online Social Networks should play in our society – and we want to take up citizen perspectives in the process of developing policy recommendations. This is why Citizen Labs play an important role in our project. In discussions between scientists from our project and citizens, we developed a draft of the first policy brief, which we then reviewed with our experts from the consortium in a workshop during a consortium meeting, as well as in a later feedback process.

At the first Citizen Lab in Karlsruhe in September 2024, participants and the public engaged in workshops, lectures, a BarCamp and a World Café session on the influence of online social networks on democracy. With input from experts, they explored topics such as the dynamics of social media, the EU's Digital Services Act and media literacy. Participants worked together to produce a policy brief on necessary regulatory action, combining their learning with practical insights. The process was guided by experts from the TWON consortium, including Prof. Dr. Damian Trilling (workshop on “The Limits of Research on Social Media Dynamics”), Prof. Dr. Achim Rettinger (workshop on “The Role of Social Media and AI in the Confluence of Real-Life Crisis and Digital Democracy - A Technical Perspective”) and Dr. Eugen Pissarskoi (workshop on “Can Two Wrongs make a Right? – An Ethical Reflection on the Idea of Creating Twins of Online Social Networks?”). The workshops also benefited from the insights of external collaborators from “ISD Institute for Strategic Dialogue” such as Melanie Döring, Project Coordinator “Digital Policy Lab”, and Marisa Wengeler, Senior Educator Business Council for Democracy, who held workshops on “The Possibilities and Limitations of the EU's Digital Services Act” as well as “Pre- & De-bunking in the Online Realm”, ensuring a well-rounded approach to the development of recommendations.

Following the Citizen Lab, the recommendations were then discussed extensively within the TWON consortium during a workshop at a consortium meeting in October 2024. By doing so, we ensured that the recommendations were comprehensive and practical and covered ideas from our diverse research fields. The workshop was supported by Judith Peterka, Member of the TWON Advisory Board and Advisor at the German Federal Chancellery for AI.

The combination of academic perspectives with the insights of DialoguePerspectives participants and the public enriched the policy recommendation process. While scientific research provides empirical and theoretical foundations, DialoguePerspectives' participatory approach ensures that the recommendations are comprehensible, linked to the ongoing public debate and reflect the lived experiences of the diverse European communities. This interplay enhances the relevance and applicability of our policy advice, which we will communicate to policymakers and industry leaders in the process of our project.

The Citizen Labs are conducted by TWON consortium member “DialoguePerspectives. Discussing Religions and Worldviews e.V.” within the framework of the DialoguePerspectives program. DialoguePerspectives trains young European leaders in the sciences, culture, politics and business to become experts in a new, society-oriented interreligious-worldview dialogue

DialoguePerspectives contributes significantly to understanding and cooperation in Europe, strengthens and defends European civil society and strives to shape a pluralistic, democratic and solidary Europe. The program brings together participants from diverse communities and backgrounds, encompassing individuals with 19 different religions and beliefs across 25 European countries. Through their unique perspectives and expertise, they contribute to fostering understanding, cooperation, and a pluralistic, democratic, and cohesive Europe.

Within its commitment to TWON, DialoguePerspectives integrates its established expertise in fostering pluralistic dialogue and combating societal polarization with a focused emphasis on digital democracy, hate speech, and disinformation. The program has a proven track record in formulating actionable calls to action and comprehensive policy briefs, as demonstrated through its European Leadership Workshops on topics such as “Plurality & Anti-Discrimination in the Workplace” and events like “Entering the Engine Room: Policy Briefs as a Means of Forging a Pluralistic Europe.” These initiatives include developing policy recommendations aimed at advancing a cohesive and pluralistic Europe, methods and skills it employs within TWON. In this context, DialoguePerspectives has prioritized educating participants on the dynamics of online platforms, the role of AI, and strategies like pre-bunking and de-bunking to combat disinformation—an essential step toward strengthening and promoting a pluralistic and democratic European society. These efforts underscore the program's expertise and capacity to develop actionable policy recommendations for TWON as a part to contribute to shaping a cohesive Europe. A podcast episode on democracy in the digital age with Dr. Jonas Fegert (FZI Forschungszentrum Informatik) was recorded on the role of platforms in European democracies.

Sources

Bojic, L. (2024). AI alignment: Assessing the global impact of recommender systems. *Futures*, 160, 103383. <https://doi.org/10.1016/j.futures.2024.103383>

European Commission (2018). Joint Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions. Action Plan against Disinformation.

House of Participation (2023). A Taxonomy for Involvement Projects. <https://hop.fzi.de/taxonomy>.

Luca, M., & Bazerman, M. H. (2020). Want to Make Better Decisions? Start Experimenting. *MIT Sloan Management Review*, 61(4), 67-73.

Ludwig, K., Grote, A., Iana, A., Alam, M., Paulheim, H., Sack, H., Weinhardt, C. & Müller, P. (2023). Divided by the algorithm? The (limited) effects of content-and sentiment-based news recommendation on affective, ideological, and perceived polarization. *Social Science Computer Review*, 41(6), 2188-2210.

Miller, C., Weir, D., Ring, S., Marsh, O., Inskip, C., & Chavana, N. P. (2023). Antisemitism on twitter before and after Elon Musk's acquisition. Institute for Strategic Dialogue.

Peppers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, 24(3), 45-77.

Srnicek, N. (2017). Platform capitalism. John Wiley & Sons.

Taylor, C., Nanz, P., & Taylor, M. B. (2020). *Reconstructing Democracy: How citizens are building from the ground up*. Harvard University Press. TWONs