Digital Democracy & Social Cohesion: From Promise to Practice

Jonas Fegert
Karlsruhe Institute of Technology & FZI Research Center for Information Technology
fegert@fzi.de

Olga Slivko
Erasmus University Rotterdam
slivko@rsm.nl

Stefan Stieglitz
University of Potsdam
stefan.stieglitz@uni-potsdam.de

Christof Weinhardt
Karlsruhe Institute of Technology
weinhardt@kit.edu

1. Introduction

For years, online social networks (OSN) were promoted as tools to bring us closer together. We embraced that promise a little too readily. In Information Systems (IS), we often put efficiency first, by focusing on how digital systems and platforms could speed up workflows, enable faster coordination, and improve performance. When in e-government studies, governmental institutions entered the picture, the story was largely one of digital transformation: transform the infrastructure to a digital one, and effectiveness, trust, and legitimacy would follow. That vision has grown thin. Even in the most "digitally advanced" societies, we now see that transformation alone cannot solve what scholars and policymakers increasingly describe as a crisis of democracy and that the very information systems we design have, in part, contributed to it (Stieglitz et al., 2025; Weinhardt et al., 2024).

Over the past decade, research across IS, computational social science, and political communication has revealed how social media mechanisms can undermine trust, amplify division, and destabilize shared facts. Analytical frameworks (Stieglitz et al., 2014; Stieglitz & Dang-Xuan, 2013), developed in IS, have contributed to finding the right analytical tools for empirical OSN research. Recent longitudinal work finds that while traditional media use is associated with higher institutional trust, heavy social media use predicts lower trust, with evidence that online exposure may precede its decline (Borukhson et al., 2025). Cross-platform analyses confirm that algorithmic curation favors emotionally charged and out-group–focused content. The type of messaging most likely to elicit engagement but least conducive to understanding (Rathje et al., 2021; Yarchi et al., 2021). Bhadani et al. (2022) demonstrate how ranking algorithms can narrow political diversity in news exposure, while Guess et al. (2023) show that even small algorithmic tweaks to news feeds can measurably shift political attitudes during elections. These effects are not uniform, as Nyhan et al. (2023) note that most users inhabit mixed information environments, but the cumulative evidence is sobering: information systems and their design matter for democratic resilience.

Such findings echo the warning that technologies and platforms are not neutral intermediaries; they could be also seen as governance systems that distribute visibility, shape deliberation, and influence trust. When their logic of engagement collides with the needs of deliberative democracy in a Habermasian sens (Habermas, 1990), the result is a widening gap between technological innovation and social cohesion.

We have learned, sometimes painfully, that efficiency is not enough. The question is no longer only how digital systems and platforms can *enable* participation, but moreover how they can *protect* it. How they can be designed to strengthen trust, transparency, and inclusion rather than erode them.

2. Reframing the IS Research Agenda

This mini-track on Digital Democracy & Social Cohesion at HICSS 2025 responds to that challenge. We aim to connect three threads of inquiry that too often run in parallel:

1. Mechanisms: Understanding how algorithms, interfaces, and affordances shape civic behavior;
2. Design and Intervention: Developing systems that foster deliberation and inclusion;
3. Governance and Evaluation: Building infrastructures that ensure accountability, auditability, and democratic oversight.

URI: https://hdl.handle.net/10125/112206
978-0-9981331-9-5

Digital platforms do not merely transmit information – they structure attention and perception. Algorithmic personalization and network clustering can either connect or divide users. When recommender systems privilege similarity, users perceive polarization (Lerman et al., 2024). The empirical and theoretical groundwork is strong. Research shows that mis- and disinformation fuel polarization (Ecker et al., 2022). At the same time, studies reveal promising design and explainable-AI approaches to mitigate those tendencies (Bezzaoui, Jarvers, et al., 2025; Bezzaoui, Stein, et al., 2025; Lasser & Poechhacker, 2025). Civic experiments with online deliberation confirm that algorithmic tools can scale participation when guided by transparent rules (Alnemr, 2024). Likewise, regulatory frameworks can directly influence the quality and inclusiveness of online discourse (Andres & Slivko, 2021). For IS research, this raises a crucial question: which design principles make social cohesion measurable and sustainable?

Through design insights can become interventions. Therefore, scholars argue that platform design can explicitly aim to promote pluralism and deliberation, aligning with the "resilient digital democracies" (Weinhardt et al., 2024), which defines democratic resilience as a design goal.

Finally, resilience requires functioning institutions. As open-source AI models proliferate, safety guardrails remain inconsistent. Evaluations show that even "aligned" models can reproduce harmful, extremist or racist content (Ganguli et al., 2022; Hofmann et al., 2024; Jaidka et al., 2025). Ensuring that such systems serve democratic rather than divisive purposes therefore remains one of the most pressing challenges for Information Systems research.

3. How the accepted papers advance this agenda

The three papers selected for this mini-track exemplify how Information Systems research can respond to the challenges outlined above: moving from diagnosis to design, governance, and evaluation of democratic infrastructures. Each contribution addresses a distinct layer of digital democracy.

(1) *Virtual Encounters, Real Impact? How Social Media Affordances Foster Intergroup Contact* (Voronin & Stieglitz): This systematic review (2012–2025; 30 studies) synthesizes evidence on how social media affordances can foster positive intergroup contact and social cohesion. The findings are cautiously optimistic: most studies document improved intergroup attitudes across ethnic, religious, and sexual divisions. The paper reframes platforms as potential bridges rather than amplifiers of division and identifies where future IS design can most effectively strengthen inclusion and cohesion.

(2) *Guardrail Vulnerabilities in Open-Source Language Models: Implications for Democratic Discourse and Marginalized Communities* (Münker & Sartori)

This empirical study probes seven open-source large language models and uncovers recurring safety failures that lead to toxic or biased outputs under adversarial prompts. Using NLP-based classification and red-teaming approaches, the authors reveal how openness without oversight can expose vulnerable communities to harm. They argue that genuine democratization of AI requires robust governance that includes transparent data documentation, pre-deployment testing, and inclusive red-team pipelines, thereby ensuring safety advances within open-source LLMs.

3. *An LLM-Based Multi-Agent System for the Political Assessment of Chat-LLMs* (Johnson & Schaal) Adopting a design science lens, this study introduces a multi-agent framework that autonomously audits political bias in chat models using the Wahl-O-Mat questionnaire from the German Federal Agency for Civic Education. The system evaluates models like Mistral-Large-2, revealing a consistent center-left tendency and generating transparent visualizations of results. Beyond its technical contribution, the paper exemplifies how AI can be repurposed as an auditor and tool for transparency and democratic accountability.

Collectively, these papers span design (creating bridging affordances), governance (exposing vulnerabilities), and evaluation (building audit mechanisms). Our hope is that this mini-track and its contributions help in cultivating an IS community that not only diagnoses problems but also builds resilient alternatives: Information systems that support societal cohesion by design. The three accepted papers are a starting point, not an endpoint, and we look forward to the conversations and collaborations they will spark at HICSS and beyond.

4. References

Alnemr, N. (2024). Deliberative democracy in an algorithmic society: Harms, contestations and deliberative capacity in the digital public sphere. *Democratization*, *0*(0), 1–20. https://doi.org/10.1080/13510347.2025.2522920

Andres, R., & Slivko, O. (2021). Combating online hate speech: The impact of legislation on Twitter. *ZEW*

Discussion Papers, Article 21–103. https://ideas.repec.org//p/zbw/zewdip/21103.html

Bezzaoui, I., Jarvers, L., Weinhardt, C., & Fegert, J. (2025). Designing Deepfake Detection Systems: Practitioner Requirements Across Sectors. *To Be Released in: Proceedings of the International Conference on Information Systems (ICIS 2025, Accepted)*.

Bezzaoui, I., Stein, C., Weinhardt, C., & Fegert, J. (2025). Explainable AI for online disinformation detection: Insights from a design science research project. *Electronic Markets*, *35*(1), 66. https://doi.org/10.1007/s12525-025-00799-3

Bhadani, S., Yamaya, S., Flammini, A., Menczer, F., Ciampaglia, G. L., & Nyhan, B. (2022). Political audience diversity and news reliability in algorithmic ranking. *Nature Human Behaviour*, *6*(4), 495–505. https://doi.org/10.1038/s41562-021-01276-5

Borukhson, D., Fegert, J., & Lorenz-Spreen, P. (2025). *Diverging associations of traditional versus social media with government trust*. https://doi.org/10.31219/osf.io/7yvx6_v1

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

Ganguli, D., Lovitt, L., Kernion, J., Askell, A., Bai, Y., Kadavath, S., Mann, B., Perez, E., Schiefer, N., Ndousse, K., Jones, A., Bowman, S., Chen, A., Conerly, T., DasSarma, N., Drain, D., Elhage, N., El-Showk, S., Fort, S., … Clark, J. (2022). *Red Teaming Language Models to Reduce Harms: Methods, Scaling Behaviors, and Lessons Learned* (No. arXiv:2209.07858). arXiv. https://doi.org/10.48550/arXiv.2209.07858

Guess, A. M., Malhotra, N., Pan, J., Barberá, P., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., González-Bailón, S., Kennedy, E., Kim, Y. M., Lazer, D., Moehler, D., Nyhan, B., Rivera, C. V., Settle, J., Thomas, D. R., … Tucker, J. A. (2023). How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science*, *381*(6656), 398–404. https://doi.org/10.1126/science.abp9364

Habermas, J. (1990). *Strukturwandel der Öffentlichkeit: Untersuchungen zu einer Kategorie der bürgerlichen Gesellschaft: mit einem Vorwort zur Neuauflage 1990* (15th ed.). Suhrkamp.

Hofmann, V., Kalluri, P. R., Jurafsky, D., & King, S. (2024). AI generates covertly racist decisions about people based on their dialect. *Nature*, *633*(8028), 147–154. https://doi.org/10.1038/s41586-024-07856-5

Jaidka, K., Chen, T., Chesterman, S., Hsu, W., Kan, M.-Y., Kankanhalli, M., Lee, M. L., Seres, G., Sim, T., Taeihagh, A., Tung, A., Xiao, X., & Yue, A. (2025). Misinformation, Disinformation, and Generative AI: Implications for Perception and Policy. *Digit. Gov.: Res. Pract.*, *6*(1), 11:1-11:15. https://doi.org/10.1145/3689372

Lasser, J., & Poechhacker, N. (2025). Designing social media content recommendation algorithms for societal good. *Annals of the New York Academy of Sciences*, *1548*(1), 20–28. https://doi.org/10.1111/nyas.15359

Lerman, K., Feldman, D., He, Z., & Rao, A. (2024). Affective polarization and dynamics of information spread in online networks. *Npj Complexity*, *1*(1), 8. https://doi.org/10.1038/s44260-024-00008-w

Nyhan, B., Settle, J., Thorson, E., Wojcieszak, M., Barberá, P., Chen, A. Y., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., González-Bailón, S., Guess, A. M., Kennedy, E., Kim, Y. M., Lazer, D., Malhotra, N., Moehler, D., … Tucker, J. A. (2023). Like-minded sources on Facebook are prevalent but not polarizing. *Nature*, *620*(7972), 137–144. https://doi.org/10.1038/s41586-023-06297-w

Rathje, S., Van Bavel, J. J., & Van Der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences*, *118*(26), e2024292118. https://doi.org/10.1073/pnas.2024292118

Stieglitz, S., & Dang-Xuan, L. (2013). Social media and political communication: A social media analytics framework. *Social Network Analysis and Mining*, *3*(4), 1277–1291. https://doi.org/10.1007/s13278-012-0079-3

Stieglitz, S., Dang-Xuan, L., Bruns, A., & Neuberger, C. (2014). Social Media Analytics. *Business & Information Systems Engineering*, *6*(2), 89–96. https://doi.org/10.1007/s12599-014-0315-7

Stieglitz, S., Fegert, J., Krasnova, H., Voronin, G., & Weinhardt, C. (2025). Social media and society. *I-Com*. https://doi.org/10.1515/icom-2025-0029

Weinhardt, C., Fegert, J., Hinz, O., & van der Aalst, W. M. P. (2024). Digital Democracy: A Wake-Up Call. *Business & Information Systems Engineering*, *66*(2), 127–134. https://doi.org/10.1007/s12599-024-00862-x

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. https://doi.org/10.1080/10584609.2020.1785067