

Expert Foresight on AI-Driven Disinformation: Findings from the ATHENA Mini-Delphi Study

<https://doi.org/10.22151/politikon.61.CON8>

Pedro PERES CAVALCANTE

Trilateral Research

pedro.perescavalcante@trilateralresearch.com

Abstract: *The 2024 “super election year” demonstrated both the promise and risks of artificial intelligence (AI) in democratic processes. Innovative uses, such as AI-powered fact-checking and voter engagement tools, contrasted with harmful applications including deepfakes. Although the overall impact of deepfakes and other AI-generated disinformation was lower than anticipated, experts predict a substantial rise in AI-powered disinformation in the short term, particularly through synthetic media and foreign influence operations. Evidence of AI-enabled manipulation highlights the need for effective regulation and strengthened oversight to safeguard information integrity in a rapidly changing information ecosystem.*

Keywords: Artificial Intelligence and Democracy; Foreign Information Manipulation and Interference (FIMI); Disinformation; Misinformation; Delphi Study

In 2024, more than half of the world’s population voted in more than sixty elections. The breadth of Artificial Intelligence (AI) use cases during this “super election year” demonstrated the technology’s potential to disrupt electoral processes—both positively and negatively.

Uses such as AI-powered preliminary fact-checking (Corney, Wilkinson, and Cann 2024) and an AI avatar collecting voters’ ideas to generate public-policy proposals (Grierson 2024) showcased AI’s innovative potential, while disinformation deepfakes that attributed false statements to political figures and deepfake nudes amplified gendered political violence (Stockwell 2024; Schneier and Sanders 2024). For some countries, 2024 marked the first elections in which generative AI was readily available. The anticipated scale of impact from disinformation deepfakes, however, was widely considered to have been overestimated by the end of last year (Barr 2025; Schneier and Sanders 2024).

Nevertheless, in the medium term, AI applications are expected to be the most impactful emerging technology to disinformation, as revealed by a recent Delphi study by the ATHENA project (Cavalcante and Wright 2025). Twenty-seven experts were surveyed on their views on the evolution of disinformation in the next three to five years, with the vast majority predicting a significant surge in disinformation content: 88% of respondents estimated an increase of between 100% and 1000%. When asked to rank the expected impact of eight emerging technologies, experts predicted a prevalence of generative-AI-fuelled disinformation; they identified deepfakes/synthetic media and LLMs as likely to play a significant role in disrupting the information space. Among the expected impacts of these emerging technologies, respondents

most often mentioned the amplification of foreign influence operations capable of destabilising countries.

While evidence on the extent of harm caused by AI-generated disinformation remains limited—as illustrated by the fake audio incident during the 2023 Slovakian elections (De Nada and Peter Jančárik 2024)—AI nevertheless presents a significant threat to information integrity more broadly (EDMO 2025). As the technology redefines online information-seeking, malicious uses — such as poisoning by foreign actors (Sadeghi and Blachez 2025)—weaken the information ecosystem and deepen the trust crisis.

While some experts advocated for stricter rules, others emphasised the need for measures that counter disinformation without compromising freedom of expression. This balanced approach is materialised in the European Democracy Shield (EUDS), the EU’s recent initiative that outlines strategies and measures to tackle disinformation and FIMI across three axes. A notable strength of the EUDS that it mobilises a broad range of stakeholders, from fact-checkers and civil society organisations to VLOPs and VLOSEs¹. Among the various actions the EUDS proposes, strengthening the co-regulatory framework, embodied in the Code of Conduct on Disinformation, stands out as one of the most difficult tasks ahead, given its reliance on voluntary commitments. While the EUDS states that the European Commission will assess the level of commitment of its signatories, establish dialogues with follow-up actions, and pursue more concrete efforts on issues such as recommender-system transparency and the demonetisation of disinformation (European Commission 2025), previous research has shown how performative the commitment reports submitted by VLOPs and VLOSEs can be (Borz et al. 2024; Münges and Park 2024). If the challenges that AI poses to FIMI and disinformation are to be effectively addressed, robust EU oversight of this co-regulatory approach will be essential.

References

- Barr, Alistair. 2024. “A Tsunami of AI Deepfakes Was Expected This Election Year. Here’s Why It Didn’t Happen.” *Business Insider*. <https://www.businessinsider.com/ai-deepfakes-election-year-truemedia-oren-etzioni-2024-12>.
- Bentzen, Naja. 2025. *European Democracy Shield*. European Parliamentary Research Service: PE 775.835. [https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/775835/EPRS_BRI\(2025\)775835_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/775835/EPRS_BRI(2025)775835_EN.pdf).

¹ VLOPs and VLOSEs are acronyms introduced in the European Union’s Digital Services Act (DSA) and refer to Very Large Online Platforms and Very Large Online Search Engines, respectively. These designations apply to services with more than 45 million monthly active users in the EU, placing them under the DSA’s most stringent obligations—such as systemic risk assessments, enhanced content moderation duties, independent audits, and increased transparency requirements.

- Borz, Gabriela, Fabrizio De Francesco, Thomas L. Montgomerie, and Michael Peter Bellis. 2024. "The EU soft regulation of digital campaigning: regulatory effectiveness through platform compliance to the code of practice on disinformation." *Policy Studies* 45, no. 5: 709-729. <https://doi.org/10.1080/01442872.2024.2302448>.
- Cavalcante, Pedro, and David Wright. 2025. *Mini-Delphi Consensus Report*. ATHENA Project. <https://project-athena.eu/wp-content/uploads/2025/07/T4.2-Consolidated-Consensus-Report-Mini-Delphi-1.pdf>.
- Corney, David, Kate Wilkinson, and Robert Cann. 2024. "The AI Election: How Full Fact Is Leveraging New Technology for UK General Election Fact Checking." *Full Fact* (blog) <https://fullfact.org/blog/2024/jun/the-ai-election-how-full-fact-is-leveraging-new-technology-for-uk-general-election-fact-checking/>.
- De Nadal, Lluís, and Peter Jančárik. 2024. "Beyond the Deepfake Hype: AI, Democracy, and 'the Slovak Case'." *HKS Misinformation Review*. <https://doi.org/10.37016/mr-2020-153>.
- EDMO (European Digital Media Observatory). 2025. *European Digital Media Observatory Fact-checking Brief (September 2025)*. Edited by Lucia Bertoldini and Tommaso Canetta. European Digital Media Observatory. https://belux-edmo.s3-accelerate.amazonaws.com/wp-content/uploads/2025/10/52-EDMO-Brief-September-2025_EN.pdf.
- European Commission. 2025. "European Democracy Shield: Empowering Strong and Resilient Democracies" Brussels, 12 November 2025. JOIN(2025) 791 final. Document 52025JC0791. <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52025JC0791>.
- Grierson, Jamie. 2024. "Brighton General Election Candidate Aims to Be UK's First 'AI MP'." *The Guardian*. <https://www.theguardian.com/politics/article/2024/jun/10/brighton-general-election-candidate-uk-first-ai-mp-artificial-intelligence>.
- Jahangir, Ramsha. 2025. "EU Disinformation Code Takes Effect Amid Censorship Claims and Trade Tensions." *Tech Policy Press*: blog. <https://www.techpolicy.press/eu-disinformation-code-takes-effect-amid-censorship-claims-and-trade-tensions/>.
- Sadeghi, McKenzie, and Isis Blachez. 2025. "A Well-Funded Moscow-Based Global 'News' Network Has Infected Western Artificial Intelligence Tools Worldwide with Russian Propaganda." *NewsGuard's Reality Check* (blog) <https://www.newsguardrealitycheck.com/p/a-well-funded-moscow-based-global>.
- Schneier, Bruce, and Nathan E. Sanders. 2024. "The Apocalypse That Wasn't: AI Was Everywhere in 2024's Elections, but Deepfakes and Misinformation Were Only Part of the Picture." *HKS Misinformation Review*. <https://www.ash.harvard.edu/articles/the-apocalypse-that-wasnt-ai-was-everywhere-in-2024s-elections-but-deepfakes-and-misinformation-were-only-part-of-the-picture/>.
- Stockwell, Sam. 2024. *AI-Enabled Influence Operations: Threat Analysis of the 2024 UK and European Elections*. The Alan Turing Institute. https://cetas.turing.ac.uk/sites/default/files/2024-09/cetas_briefing_paper-ai-enabled_influence_operations-threat_analysis_of_the_2024_uk_and_european_elections.pdf