

Sinanoglu, Semuhi; Breuer, Anita

Research Report

Countering information pollution to protect democracy

IDOS Policy Brief, No. 29/2024

Provided in Cooperation with:

German Institute of Development and Sustainability (IDOS), Bonn

Suggested Citation: Sinanoglu, Semuhi; Breuer, Anita (2024) : Countering information pollution to protect democracy, IDOS Policy Brief, No. 29/2024, German Institute of Development and Sustainability (IDOS), Bonn

This Version is available at:

<https://hdl.handle.net/10419/302802>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>



Countering Information Pollution to Protect Democracy

Semuhi Sinanoglu & Anita Breuer

Summary

Information pollution poses significant challenges to democracy by undermining informed decision-making and threatening social cohesion. The increasing sophistication and accessibility of artificial intelligence (AI), including the generation of deep fakes, exacerbates these challenges by making it harder to discern truth from falsehood, thus manipulating public perception. In addition, information pollution disproportionately affects disadvantaged groups, sometimes inciting online and offline violence. Although an extensive toolkit of interventions to counter information pollution exists, the current debate is narrowly focused on content regulation.

Several international initiatives have emerged aiming to safeguard information integrity in the digital space, particularly during elections. However, diverse national approaches to data governance and the ongoing worldwide trend of autocratisation make it challenging to agree on a global, principled position against information pollution.

Further, cross-sectoral cooperation in this area needs to be improved. Increased collaboration with technology companies and the private sector will be indispensable to developing and implementing an international regulatory framework. The Global Digital Compact, which will be negotiated by UN member states at the Summit for the Future from 22 to 23 September, aims to deepen cooperation and establish a global framework for a digital future.

This policy brief presents international initiatives and critically discusses the tools available to combat information pollution. Our recommendations are as follows:

- **Multilateral and cross-sectoral cooperation will be key to establishing a transnational regulatory framework for a safe and inclusive digital space;** this requires the active engagement of all relevant stakeholders, including the private sector, in digital cooperation.
- **The focus of approaches to counter information pollution needs to be broadened.** Instead of narrowly focusing on politically controversial content moderation, content-neutral intervention tools should receive more attention, and the potential of AI tools to scale up these interventions should be considered.
- **Policy-makers and advisors need to be mindful of context sensitivity.** There is no one-size-fits-all solution, and effective tools in some contexts may backfire in others. Thus, interventions to counter information pollution should only be integrated into national policies or development programmes with thorough prior research.
- **There are no quick fixes.** Long-term, holistic approaches are required to enhance societies' resilience against information pollution. Such strategies must go beyond the digital realm, including effective measures to support independent media and free flow of information, as well as the inclusion of information literacy into school curricula and educational programmes.

Digital challenges to democracy and structural challenges for international and development cooperation

The rising tide of information pollution poses significant challenges to democracies and societal peace worldwide. Democracy fundamentally relies on an informed public, freedom of expression, media independence and equal access to information. Accurate and unbiased information enables citizens to make sound political and economic decisions. However, citizens' exposure to misleading and contradictory information about political issues on digital platforms can increase polarisation and decrease trust in democratic institutions and the truthfulness of information in general.

The issue of information pollution is particularly acute during elections when information manipulation is often employed to discredit political opponents, critical journalists and activists or undermine trust in electoral authorities and processes.

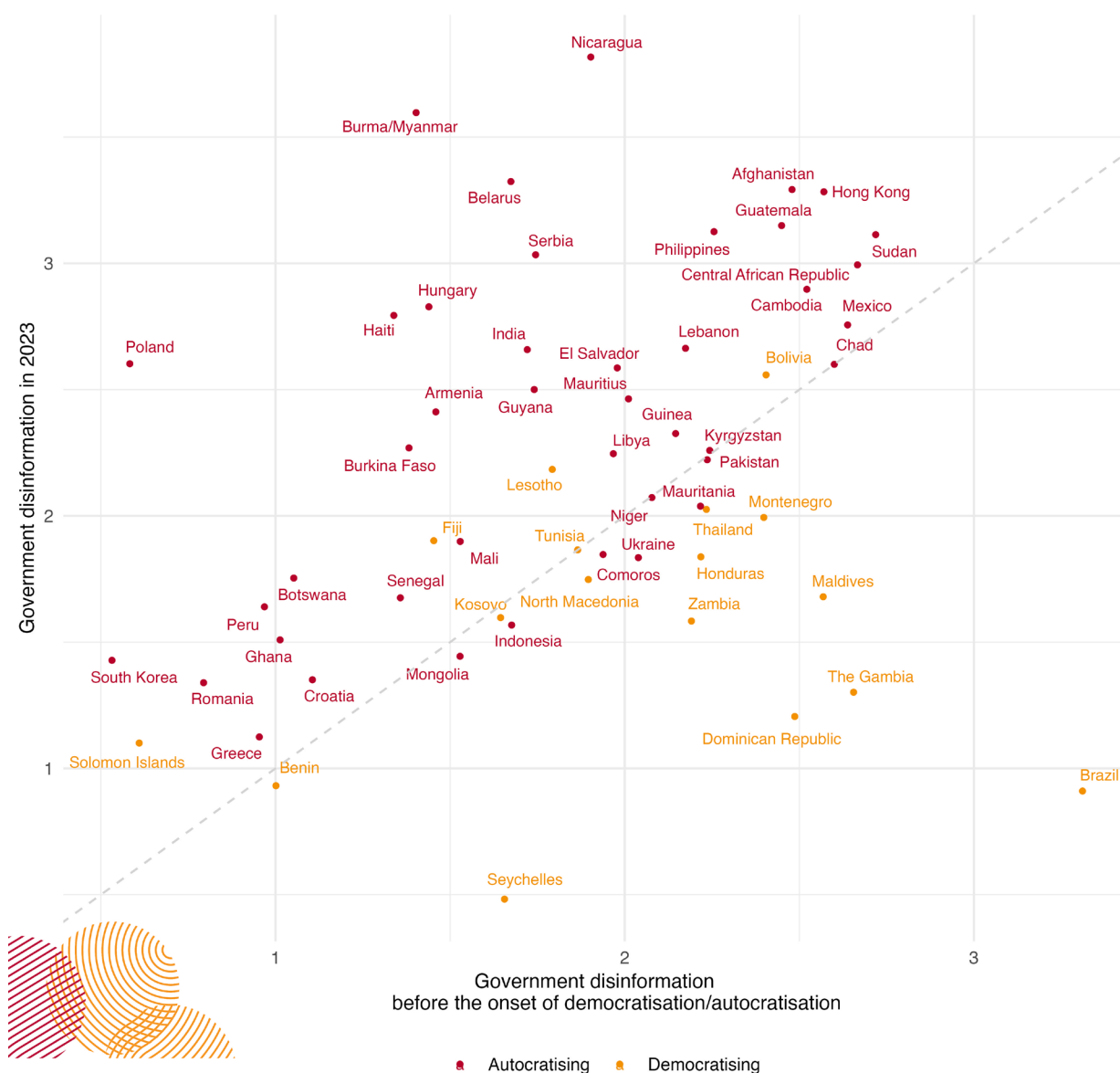
Attacks on information integrity are a hallmark of authoritarian political control – autocrats strategically use disinformation to maintain and expand their power. Figure 1 illustrates the relationship between autocratisation and government disinformation. The vertical axis displays the levels of disinformation disseminated by governments domestically. The horizontal axis displays the disinformation scores before the onset of democratisation or autocratisation episodes, which started in different years for different countries. Only countries that experienced a regime transformation are included and marked either “auto-cratising” or “democratising” in 2023. Although merely descriptive, the graph suggests that auto-cratising countries exhibit higher levels of government disinformation.

Box 1: Information manipulation and pollution

A standardised terminology related to the dissemination of manipulated or low-quality information has not yet evolved. Popular terms like “fake news” are narrowly focused and have been co-opted by political actors who use them to delegitimise critical media reporting. By contrast, the broader and value-neutral concept of information pollution (UNDP, 2022) encompasses different types of low-quality information in the information ecosystem that differ in terms of intent and strategies for dissemination:

Misinformation refers to false or inaccurate content that is shared without the intention to cause harm to a person, group or country. **Malinformation** refers to a situation where genuine information is intentionally manipulated and shared to inflict harm. **Disinformation** means false information intentionally created to cause harm and disseminated using practices that go well beyond news reporting (such as automated accounts, targeted advertising, organised online trolling and internet memes).

Figure 1: The amount of disinformation disseminated by governments domestically and political regime transformation (autocratisation or democratisation)



Source: Reproduced from V-DEM (2024)

Mexico constitutes an illustrative case in point. V-DEM's Digital Society Project registered a significant increase in government-disseminated disinformation under President Lopez Obrador (2018-2024). He actively engaged in disseminating malinformation and disinformation, frequently aimed at discrediting journalists and tarnishing their professional reputations (Breuer, 2024). These trends coincide with a decline in the country's quality of democracy. According to V-DEM data, Mexico began an autocratisation episode in 2020.

Rapid breakthroughs in artificial intelligence (AI) technologies add to digital challenges to democracy. People may become more prone to manipulation as it is challenging to distinguish AI-generated text from human-generated content (Kreps et al., 2022). AI-generated content, often designed to trigger highly emotional responses, can be hard to detect and spread rapidly across algorithm-driven platforms and media outlets. Even though generative AI is still in its infancy (Garimella & Chauchard, 2024), deep fakes will

likely be increasingly employed to disseminate manipulated information.

Further, information pollution exacerbates discrimination against disadvantaged groups and minorities, who are more vulnerable to online violence. For instance, according to a global study by UN Women (2022), 38 per cent of women have experienced digital violence. Online hate speech also translates into offline violence. Attacks on minorities by vigilante groups are often organised on digital platforms.

Against this background, global concerns about the impact of information pollution in fuelling autocratisation and polarisation are growing. Reflecting these concerns and partly motivated by the “Super Election Year 2024”, the Global Risks Report 2024 of the World Economic Forum (2024) identified misinformation and disinformation as the most severe short-term risk the world currently faces.

Challenges for international collaboration and development cooperation

In response, the international community increasingly seeks to strengthen the integrity of the global information space, as illustrated by the UN Global Principles for Information Integrity launched in June this year (UN, 2023). Active efforts to curb information pollution are also evident in both international policy debates and development cooperation.

The OECD, for example, recently reformed its principles for media support. Comprehensive consultations, led by journalists and media development organisations, concluded that the previous (2014) OECD Principles on media assistance were no longer fit for purpose given increasing global levels of information manipulation, polarisation and autocratisation. In March 2024, the OECD Development Assistance Committee Network on Governance (OECD DAC-GovNet) adopted the reformed “Principles for Relevant and Effective Support to Media and the Information Environment” (OECD, 2024). Targeting development

agencies within the DAC, as well as media support practitioners, political parties, international policy-makers, private foundations and investors, the principles advocate a strategic, holistic approach to balance the need to combat information manipulation with the protection of freedom of expression in the context of evolving technologies like AI.

Another example is the working group “Media and Digital” within the Team Europe Democracy (TED) Initiative. The group, which has the mandate to promote inclusive democracy and pluralistic, independent media as critical objectives under the European Union (EU) Strategic Agenda 2019-2024, includes representatives of EU donor agencies as well as academics, journalists and organisations that promote rights to access to information and freedom of expression. In the run-up to the UN Summit for the Future, members of the group shared recommendations on steps that EU member states and the European Commission could take to ensure that the UN Pact for the Future includes robust commitments to access to information, media freedom and public-interest journalism.

Despite these commendable initiatives, international cooperation to counter disinformation faces major challenges:

- *Structural challenge.* The increasing trend of autocratisation described above is a significant obstacle to international cooperation for information integrity. Forty-two countries are currently experiencing periods of autocratisation, with half the world under authoritarian rule. As the major propagators of manipulated information, autocrats and populists are not invested in joining a global, principled position against information pollution.
- *Funding challenge.* International development cooperation’s efforts to counter information pollution have remained limited, particularly when measured by Official Development Assistance (ODA) spending: currently, 0.3 per cent of the OECD’s ODA is dedicated to supporting media and the free flow of information.

- *Sectoral silos challenge.* International cooperation also lacks cross-sectoral engagements. Although there are some promising initiatives and growing awareness among business leaders about information pollution, the lack of coordination with the business sector will cripple any collective effort to fight information manipulation. First, businesses may unknowingly fund disinformation campaigns around major global events with their digital marketing strategies on major social media platforms because purchasing is often automated. Advertisers are often unaware of which websites their ads appear on and inadvertently fund disinformation (Ahmad et al., 2024). Secondly, digital platforms must be recognised and included as stakeholders in any multilateral initiative against the spread of manipulated information. The design-based countermeasures against information pollution require their initiative and cooperation. Examples worth following include WhatsApp's "forwarded" message label and YouTube's algorithm to feature trusted sources.
- *Tunnel-vision challenge.* The current debate on combatting information pollution is often narrowly focused on content regulation. As awareness of disinformation's risks to the democratic process grows, calls for rapid crisis response solutions are becoming louder. However, this approach is both shortsighted and too narrowly conceived.

Toolkit of interventions to counter information pollution

The public demand for content moderation is limited and depends on the topic. Even though it is widely used on several digital platforms, it poses moral dilemmas and faces several limitations. First, it is only reactive by design. It also requires the willingness and mobilisation of resources on the part of the digital platform, which should not be taken for granted. Some critics also perceive it as an assault on freedom of speech. The existing research shows that digital exposure to incivility or

intolerance does not necessarily translate to public support for removal of harmful content (Pradel et al., 2024). Instead, people may prioritise the removal of false information over the protection of freedom of speech only if it includes violent threats against minority groups.

Account suspensions are not publicly popular either. De-platforming the traffickers of polluted information is a widely-used tactic. Twitter, for example, used it following the 6 January 2021 US Capitol attack (McCabe et al., 2024). The intervention not only hampered the spread of false information, it also caused traffickers of harmful content who were not yet de-platformed to follow others in leaving Twitter. The tactic remains politically controversial. It elicits public criticism that tech companies police and censor their platforms, and there is less public support for suspending accounts than removing harmful posts (Kozyreva et al., 2023).

The range of options for protecting the information space is much broader than mere content regulation or account suspension. As a remedy for the tunnel-vision challenge, an extensive toolkit of interventions against information pollution exists (Kozyreva et al., 2024). Table 1 below provides an overview of available tools and describes their varying effectiveness and scalability.

Each tool has advantages and disadvantages to consider before applying. Even though alternative solutions are presented, their effects are context-dependent and may wear off over time; there may also be negative spillovers and scalability issues.

These alternative intervention tools can be summarised as follows:

- Tactics like *debunking* and *fact-checking* seek to contain the spread of misinformation ex-post by presenting facts and logically explaining why a piece of information is inaccurate or misleading. There is a large ecosystem of fact-checkers worldwide with established rules of conduct and methodology (EFCSN, 2024).

Table 1: An overview of the toolkit for combatting information pollution

Tool	Objective	Effectiveness and/or scalability	Issues
Content moderation and deplatforming	Removal of false or misleading content Suspending accounts	Effective in containing the spread in the short run but requires the platform's resource mobilisation	May not be publicly popular and perceived as censorship Reactive only
Debunking and fact-checking	Presenting facts and logical explanations as to why a piece of information is false or misleading	Effective in containing the circulation in the short run but resource-intensive	Decreases trust in the media and belief in credible information Reactive only May lose efficacy over time Topic-specific intervention
Accuracy-nudging	Reminding people to think about accuracy of a post/headline	Highly scalable with platform's collaboration Small effect size for the short run	May lose efficacy over time Might only work in specific contexts or for specific partisan groups or political issues
Pre-bunking	Educating users about how misinformation spreads	Robust effect size Pre-emptive Moderate scalability	Requires booster interventions Effectiveness may vary by the medium

Source: Adapted and revised from Kozyreva et al. (2024)

- Nudging tools like *accuracy prompts* aim to shift people's online behaviour by highlighting the importance of accuracy and asking people, to the best of their knowledge, whether a post/headline is accurate. The underlying premise is that people do not necessarily prefer partisanship over accuracy. When made attentive to the importance of accuracy, they become more likely to share actual content. In 2021, for example, Twitter launched the Community Notes feature allowing contributors to collaborate in registering contextual information on tweets they deem misleading.
- *Pre-bunking* – an educational intervention – aims to enhance people's ability to recognise fake news pre-emptively. Recently, for example, Google and Jigsaw launched a pre-bunking campaign ahead of the European Parliament elections to raise awareness about how malicious actors deploy fake news online. Some of these interventions include game-based solutions. For instance, millions of users

have played the online *Bad News Game*, assuming the role of a misinformation tycoon (Iyengar et al., 2023).

There is ample evidence regarding the effectiveness of this alternative toolkit. Debunking and fact-checking successfully reduce the circulation of inaccurate information, especially during mass disinformation campaigns (Unver, 2020). Simple accuracy reminders may also significantly reduce misinformation sharing on social media – people are more likely to share true headlines than false ones after having rated their accuracy (Pennycook et al., 2021). In addition, pre-bunking may have a high potential to “vaccinate” people against manipulated information (McPhedran et al., 2023).

However, there are certain contextual and design limitations to their effectiveness:

- Debunking and fact-checking are reactive by design, and their effects may be short-lived. Research shows that fact-checking does not

necessarily lead to overall decreased engagement with misinformation content (Carey et al., 2022).

- The positive effects of some accuracy nudges may be modest, decrease over time and work for specific partisan groups or contexts with low misinformation levels (Butler et al., 2024). There may also be design-related issues: even though some evidence shows that Community Notes effectively reduce the number of shares of false information (Renault et al., 2024), their immediate impact on the circulation of polluted information may be insufficient due to the time delays between their publication and the original post. The effect might also depend on the medium of intervention. Bowles et al. (2023) find that a quick WhatsApp text message that encouraged people to engage in fact-checking regularly was more effective than long podcasts.
- One limitation of educational interventions like pre-bunking is that their effects might decay if not reinforced with “booster” treatments (Maertens et al., 2021).

There may also be unintended consequences and negative spillovers. For instance, fact-checking may lower trust in the media and belief in the credibility of factual information (Hoes et al., 2024). In other words, even though fact-checking is helpful during mass disinformation campaigns, it may inadvertently increase scepticism of accurate information (Altay et al., 2023). That is why fact-checking may risk playing right into malicious actors’ hands in the long run since disinformation campaigns often aim to sow confusion and mistrust rather than political persuasion.

There is considerable variation in the scalability of these tools. On the one hand, by design, fact-checking and debunking are topic-specific with limited generalisability. They require large resources with the set industry standards. Some educational interventions and tools may not be easily scalable either. On the other hand, accuracy nudges can be easily scalable. For instance, one such scalable idea is to add a

“misleading count” next to the “like” count, allowing users to mark posts as misinformation. Higher “misleading” counts would dissuade people from circulating the post (Pretus et al., 2024).

In summary, the broader toolkit must be considered to fight information pollution. Designing scalable digital infrastructure with content-neutral interventions that combat information pollution pre-emptively and in real time is possible. There are design features and ideas, some of which have already been employed with modest success. In addition, pre-bunking offers a promising alternative as a potential pre-emptive solution to create resilience against information pollution.

However, there are no one-size-fits-all solutions to information pollution. The effectiveness of each strategy is context dependent. Policy-makers and advisors must tread carefully before translating such interventions into policy frameworks and consider contextual nuances and long-term effects.

As a remedy for scalability challenges, the role of AI in countering information pollution should be acknowledged. Even though AI exacerbates the risk of disinformation, it also presents opportunities to fight the spread of harmful content and scale up educational interventions. One recent study, for example, found that dialogues with ChatGPT may be used to debunk conspiracy theories among a group of conspiracy believers in a robust, durable and scalable way (Costello et al., 2024).

Developing long-term sustainable strategies should extend beyond the isolated application of solutions in the digital realm. Media and internet literacy initiatives and policies should be included in education curricula, ensuring the inclusion of disadvantaged youth with lower levels of education as target beneficiaries. Bearing in mind the continued digital divide, with over half the global population still lacking access to high-speed broadband, online communication campaigns to debunk false or harmful content, particularly disinformation directed against vulnerable minorities, need to be coupled with offline community-based

dialogue and awareness. Furthermore, long-term, preventive strategies also need to encourage public trust in and access to official information sources and support the media to tackle information pollution effectively. Rather than stand-alone digital interventions, efforts to counter information manipulation should be systematically integrated into thematic programming such as health, climate action, electoral and media support and the prevention of violent extremism.

Conclusions and recommendations

Information pollution poses significant challenges to democracies and social cohesion. It exacerbates polarisation, decreases trust in democratic institutions and entrenches autocratic regimes in power.

In response, public debate and international co-operation should go beyond politically divisive content regulation and take full advantage of the broader array of instruments and adapt long-term strategies to specific contexts. Some of these tactics, such as accuracy nudges and pre-bunking, offer viable alternatives, even though there are contextual limitations to their effectiveness and scalability.

To break sectoral silos and improve international funding on countering information pollution, major digital companies and businesses should be on board. Their inclusion would make applying the intervention toolkit feasible and more effective. For instance, Google initiated a major pre-bunking campaign on the eve of the European Parliament elections and also committed to funding think-tank and civil society partnerships across Central and Eastern Europe to build up research and programmes to increase media literacy (Green, 2022). Despite this growing awareness, there is still not enough cross-sectoral collaboration with businesses whose active involvement is required for a transnational regulatory framework to be created to protect information integrity.

The Summit for the Future on 22 to 23 September presents an opportunity for developing such a transnational regulatory framework to address these challenges through a long-term horizon. The draft document for the Global Digital Compact (UN, 2024), released in April, was prepared in several rounds of consultations involving technology stakeholders and will be annexed to the Summit's outcome document, the Pact for the Future, if inter-governmentally agreed. The document aims to establish international corporate accountability standards and urges companies to uphold human rights online, integrate human rights laws into emerging technologies and mitigate AI risks. Among other provisions, it calls upon digital technology companies and developers to co-create industry accountability frameworks (Art. 29b) and actions to foster an inclusive, open, safe and secure digital space (Art. 59). It is crucial that negotiating partners at the Summit for the Future ensure that these provisions remain intact to incentivise tech companies to take actions towards fostering safe and inclusive information space that go beyond occasional content moderation.

Four recommendations follow:

- **Utilise a broad toolkit of digital tactics to fight information pollution.** The global debate's focus should shift from politically divisive content moderation to other potential tactics that require less active-human moderation with content-neutral interventions and combat information pollution pre-emptively and in real time.
- **The effectiveness of these tools and tactics must be carefully designed and tested.** There is no one-size-fits-all fix to information pollution. Even though some tools look more promising than others, such as pre-bunking, their effectiveness may wear off over time and be context-dependent. Policy-makers and advisors must tread carefully before they adapt these interventions into policy frameworks. Scalability is also a concern. AI tools may be deployed to scale up such educational interventions.

- **National governments and development cooperation partners should fund long-term strategies to enhance people's competence in discerning harmful and misleading content.** Offline literacy initiatives and community-based dialogue programmes should be introduced to combat digital and offline violence and build resilience against the repercussions of disinformation campaigns.
- **A transnational regulatory framework is required to ensure information integrity. For such cooperation to work, businesses must be included as stakeholders.** Multilateral initiatives for combatting information pollution require the businesses' active collaboration. A few large technology companies significantly influence global information flows and control digital experiences. A transnational regulatory framework prioritising transparency and independent oversight is needed to address this power imbalance. Developing such a framework will require multilateral and multisectoral cooperation. The Global Digital Compact, which will be negotiated at the Summit of the Future in September, offers a unique opportunity to lay the foundations for such cooperation.

References

- Ahmad, W., Sen, A., Eesley, C., & Brynjolfsson, E. (2024). Companies inadvertently fund online misinformation despite consumer backlash. *Nature*, 630(8015), 123-131. <https://doi.org/10.1038/s41586-024-07404-1>
- Altay, S., Lyons, B., & Modirrousta-Galian, A. (2023). *Exposure to higher rates of false news erodes media trust and fuels overconfidence*. OSF Preprint. <https://doi.org/10.31234/osf.io/t9r43>
- Bowles, J., Croke, K., Larreguy, H., Marshall, J., & Liu, S. (2023). *Sustaining exposure to fact-checks: misinformation discernment, media consumption, and its political implications* (SSRN Scholarly Paper 4582703). <https://doi.org/10.2139/ssrn.4582703>
- Breuer, A. (2024). *Information integrity and information pollution: Vulnerabilities and impact on social cohesion and democracy in Mexico* (Discussion Paper 2/2024). German Institute of Development and Sustainability (IDOS). <https://www.idos-research.de/discussion-paper/article/information-integrity-and-information-pollution-vulnerabilities-and-impact-on-social-cohesion-and-democracy-in-mexico/>
- Butler, L. H., Prike, T., & Ecker, U. K. H. (2024). Nudge-based misinformation interventions are effective in information environments with low misinformation prevalence. *Scientific Reports*, 14(1), 1-12. <https://doi.org/10.1038/s41598-024-62286-7>
- Carey, J. M., Guess, A. M., Loewen, P. J., Merkley, E., Nyhan, B., Phillips, J. B., & Reifler, J. (2022). The ephemeral effects of fact-checks on COVID-19 misperceptions in the United States, Great Britain and Canada. *Nature Human Behaviour*, 6(2), 236-243. <https://doi.org/10.1038/s41562-021-01278-3>
- Costello, T. H., Pennycook, G., & Rand, D. (2024). *Durably reducing conspiracy beliefs through dialogues with AI*. OSF. <https://doi.org/10.31234/osf.io/xcwdn>
- EFCSN (European Fact-Checking Standards Network). (2024). Code of standards. *European Fact-Checking Standards Network*. <https://efcsn.com/code-of-standards/>
- Garimella, K., & Chauchard, S. (2024). How prevalent is AI misinformation? What our studies in India show so far. *Nature*, 630(8015), 32-34. <https://doi.org/10.1038/d41586-024-01588-2>
- Green, Y. (2022, 12 December). Disinformation as a weapon of war: The case for prebunking. *Friends of Europe*. <https://www.friendsofeurope.org/insights/disinformation-as-a-weapon-of-war-the-case-for-prebunking/>
- Hoes, E., Aitken, B., Zhang, J., Gackowski, T., & Wojcieszak, M. (2024). Prominent misinformation interventions reduce misperceptions but increase scepticism. *Nature Human Behaviour*, 1-9. <https://doi.org/10.1038/s41562-024-01884-x>
- Iyengar, A., Gupta, P., & Priya, N. (2023). Inoculation against conspiracy theories: A consumer side approach to India's fake news problem. *Applied Cognitive Psychology*, 37(2), 290-303. <https://doi.org/10.1002/acp.3995>
- Kozyreva, A., Herzog, S. M., Lewandowsky, S., Hertwig, R., Lorenz-Spreen, P., Leiser, M., & Reifler, J. (2023). Resolving content moderation dilemmas between free speech and harmful misinformation. *Proceedings of the National Academy of Sciences*, 120(7), 1-12. <https://doi.org/10.1073/pnas.2210666120>
- Kozyreva, A., Lorenz-Spreen, P., Herzog, S. M., Ecker, U. K. H., Lewandowsky, S., Hertwig, R., Ali, A., Bak-Coleman, J., . . . Wineburg, S. (2024). Toolbox of individual-level interventions against online misinformation. *Nature Human Behaviour*, 1-9. <https://doi.org/10.1038/s41562-024-01881-0>
- Kreps, S., McCain, R. M., & Brundage, M. (2022). All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. *Journal of Experimental Political Science*, 9(1), 104-117. <https://doi.org/10.1017/XPS.2020.37>
- Maertens, R., Roozenbeek, J., Basol, M., & van der Linden, S. (2021). Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments. *Journal of Experimental Psychology: Applied*, 27(1), 1-16. <https://doi.org/10.1037/xap0000315>
- McCabe, S. D., Ferrari, D., Green, J., Lazer, D. M. J., & Esterling, K. M. (2024). Post-January 6th deplatforming reduced the reach of misinformation on Twitter. *Nature*, 630(8015), 132-140. <https://doi.org/10.1038/s41586-024-07524-8>

- McPhedran, R., Ratajczak, M., Mawby, M., King, E., Yang, Y., & Gold, N. (2023). Psychological inoculation protects against the social media infodemic. *Scientific Reports*, 13(1), 5780. <https://doi.org/10.1038/s41598-023-32962-1>
- Mechkova, V., Pemstein, D., Seim, B., & Wilson, Steven. L. (2024). Measuring online political activity: Introducing the digital society project dataset. *Journal of Information Technology & Politics*, 1-17. <https://doi.org/10.1080/19331681.2024.2350495>
- OECD (Organisation for Economic Co-operation and Development). (2024). *Development co-operation principles for relevant and effective support to media and the information environment*. https://www.oecd-ilibrary.org/development/development-co-operation-principles-for-relevant-and-effective-support-to-media-and-the-information-environment_76d82856-en
- Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, 592(7855), 590-595. <https://doi.org/10.1038/s41586-021-03344-2>
- Pradel, F., Zilinsky, J., Kosmidis, S., & Theocharis, Y. (2024). Toxic speech and limited demand for content moderation on social media. *American Political Science Review*, 1-18. <https://doi.org/10.1017/S000305542300134X>
- Pretus, C., Javeed, A. M., Hughes, D., Hackenburg, K., Tsakiris, M., Vilarroya, O., & Van Bavel, J. J. (2024). The Misleading count: An identity-based intervention to counter partisan misinformation sharing. *Phil. Trans. R. Soc. B*, 379(1897), 1-9. <https://doi.org/10.1098/rstb.2023.0040>
- Renault, T., Amariles, D. R., & Troussel, A. (2024). *Collaboratively adding context to social media posts reduces the sharing of false news* (arXiv:2404.02803). arXiv. <https://doi.org/10.48550/arXiv.2404.02803>
- UN (United Nations). (2023). *Information integrity on digital platforms* (Our Common Agenda Policy Brief 8, pp. 1-28). <https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-information-integrity-en.pdf>
- UN. (2024, 10 September). *GDC Rev 3 - Draft under silence procedure*. un.org: https://www.un.org/techenvoy/sites/www.un.org.techenvoy/files/general/GDC_Rev_3_silence_procedure.pdf
- UNDP (United Nations Development Programme). (2022). *Information integrity: Forging a pathway to truth, resilience and trust*. <https://www.undp.org/publications/information-integrity-forging-pathway-truth-resilience-and-trust>
- UN Women. (2022). *Accelerating efforts to tackle online and technology-facilitated violence against women and girls*. <https://shknowledgehub.unwomen.org/en/resources/accelerating-efforts-tackle-online-and-technology-facilitated-violence-against-women-and>
- Unver, A. (2020). *Fact-checkers and fact-checking in Turkey* (Cyber Governance and Digital Democracy). EDAM. <https://edam.org.tr/en/cyber-governance-digital-democracy/fact-checkers-and-fact-checking-in-turkey>
- V-Dem Institute. (2024). *Democracy Report 2024: Democracy winning and losing at the ballot*. University of Gothenburg. https://v-dem.net/documents/43/v-dem_dr2024_lowres.pdf
- World Economic Forum. (2024). *Global risks report 2024*. <https://www.weforum.org/publications/global-risks-report-2024>

Dr Semuhi Sinanoglu is a researcher in the research programme “Transformation of Political (Dis-)order” at the German Institute of Development and Sustainability (IDOS).

Email: semuhi.sinanoglu@idos-research.de

Dr Anita Breuer is a project lead and senior researcher in the research programme “Transformation of Political (Dis-)order” at the German Institute of Development and Sustainability (IDOS).

Email: anita.breuer@idos-research.de

Published with financial support from the Federal Ministry for Economic Cooperation and Development (BMZ).

Suggested citation:

Sinanoglu, S., & Breuer, A. (2024). *Countering information pollution to protect democracy* (Policy Brief 29/2024). IDOS. <https://doi.org/10.23661/ipb29.2024>.

Disclaimer:

The views expressed in this paper are those of the author(s) and do not necessarily reflect the views or policies of the German Institute of Development and Sustainability (IDOS).



Except otherwise noted, this publication is licensed under Creative Commons Attribution (CC BY 4.0). You are free to copy, communicate and adapt this work, as long as you attribute the German Institute of Development and Sustainability (IDOS) gGmbH and the author(s).

IDOS Policy Brief / German Institute of Development and Sustainability (IDOS) gGmbH

ISSN (Print) 2751-4455

ISSN (Online) 2751-4463

DOI: <https://doi.org/10.23661/ipb29.2024>

© German Institute of Development and Sustainability (IDOS) gGmbH

Tulpenfeld 6, 53113 Bonn

Email: publications@idos-research.de

<https://www.idos-research.de>

Printed on eco-friendly, certified paper.

