IMPROVING EXISTING EU POLICY TO PRESERVE COGNITIVE AGENCY

Democracy in AI-Mediated Environments

Policy Brief 1 of 2

9 July 2026

Center for Environmental
and Technology Ethics · Prague

This is the first policy brief in a series on “Democracy in AI-Mediated Information Environments” and focuses on improving existing legislative frameworks, namely the DSA, the AI Act, and the Democracy Shield Initiative.

EXECUTIVE SUMMARY

The problem that social media and generative artificial intelligence (genAI) present for democracy is not limited to individual instances of mis/disinformation. They are driving a structural transformation of online information environments, making it increasingly difficult for citizens to determine which sources are credible, which claims are trustworthy, and whether content is human-generated or synthetic.
This transformation is undermining cognitive agency: namely, citizens' ability to assess and revise beliefs independently. Without it, democratic deliberation and self-determination falter. Existing EU regulation does not adequately address this structural transformation.

We recommend three policy interventions:

Expand the definitions of systemic risk and manipulation in the AI Act and the DSA to capture harms from synthetic content to information environments.

Strengthen cognitive agency by expanding digital literacy programs in the Democracy Shield Initiative.

Improve DSA enforcement to protect information environments through better research access and resources.

How Social Media and Generative AI Threaten Democracy

Citizens increasingly consume news and form political opinions on social media platforms, where opaque recommendations mediate access to information. These platforms are also flooded with AI-generated content, including text, images, videos, and what has become known as 'AI slop'. Together, these developments change not only what citizens see online, but the conditions under which they evaluate the truth and reliability of what they see.

The problem goes beyond polluted information environments. The convergence of social media and genAI poses structural threats to democracy. Democracy depends on citizens being able to form, assess, and revise their beliefs about matters of common concern; what we call **cognitive agency**. AI-driven systems increasingly shape the conditions for that agency by influencing how citizens evaluate political information and how knowledge is produced and researched within the sciences. When platform design and genAI mediate these processes, democratic values are under threat.

Existing policy instruments — the Digital Services Act (DSA), the AI Act, and the Democracy Shield — provide an important foundation (see **Textbox 1**), but require strengthening. The Commission has recognised this and promises to bring forward a Digital Fairness Act by the end of this year. A companion brief will address how a Digital Fairness Act could further help protect democracy. **This policy brief proposes three comple-mentary measures to existing legislation**. First, recognise the degradation of in-formation environments caused by synthetic content as a systemic risk and broaden existing definitions of manipulation to capture genAI-enabled forms of influence. Second, strengthen cognitive agency within the Democracy Shield by cultivating cri-tical AI literacy and epistemic literacy. Third, increase transparency through stronger re-searcher access, independent auditing, and improved enforcement of the DSA.

Three Harms to Informed Citizenship

Much of the current policy debate focuses on harmful content, misinformation, and disinformation. These concerns are important, but they do not capture the full challenge. Opaque recommender algorithms, genAI systems, and synthetic content also distort the information environments on which cognitive agency and democratic participation depend.

Harms from Synthetic Content and Manipulation

Some online content is clearly harmful, such as dangerous health advice, while other content, such as incitement to hatred or violence, is illegal under EU law. The presence of such content has been a major concern for European policymakers and ultimately contributed to the development of the DSA. Existing EU regulations already target certain manipulative practices such as: deceptive advertising, fake accounts, bot networks, and manipulative design (Art. 25, Art. 26, Art. 34 DSA; Art. 5 AI Act).

Online manipulation is often difficult to identify in practice. It can take many forms, from computational propaganda and dark patterns to nudging and some forms of persuasion. When deployed in information environments, they can hinder citizens' decision-making, weaken their ability to evaluate information critically, erode trust in democratic institutions, and undermine the capacity to meaningfully participate in those processes. In addition to those challenges, genAI introduces highly personalized and convincing but misleading content that can now be produced at scale, making manipulation harder to identify.

Detection is difficult because both the generation and the consumption of synthetic content are highly individualized: no two prompts generate the same response, and users may encounter sycophantic responses or highly biased depictions. The problem becomes particularly serious when synthetic content enters established knowledge systems, as illustrated by Google Image results for "baby peacock," where AI-generated images appear alongside photographic ones, making an accurate identification harder for unfamiliar users (see **Figure 1)**.

Textbox 1. Overview of relevant EU regulation

The **Digital Services Act (DSA)**, in effect since February 2024, regulates digital services, including digital platforms and search engines. It requires platforms, especially very large online platforms (VLOPs) such as Facebook, Instagram, TikTok, and X, to assess and mitigate systemic risks, including illegal content, discrimination, and harms to media freedom. VLOPs must also provide researcher access for independent scrutiny of their risk management systems.

The **Artificial Intelligence Act (AI Act)**, in force since August 2024, regulates AI systems according to risk level. It aims to protect fundamental rights, improve safety, and support the European Single Market through accountability and transparency obligations. It also bans certain practices, including social scoring, exploitation of group vulnerabilities, and subliminal, manipulative, or deceptive techniques that materially distort behaviour.

The **European Democracy Shield** is a policy initiative launched in November 2025 organised around three pillars: safeguarding the integrity of the information space, including by strengthening the DSA and supporting European fact-checking organisations; strengthening institutions for free and fair elections and independent media, including through guidance on the safety of political actors and support for independent journalism; and building societal resilience and citizen engagement through civic tech and evidence-based policymaking.

Harms to Cognitive Agency

For citizens to become informed, they must be able to exercise their cognitive agency: meaningful control over the intellectual capacities by which they form, assess, and revise beliefs. Cognitive agency is not simply an individual capacity. It is shaped by the socio-technical environments in which information is encountered and shared (see **Textbox 2**). The problem is not merely that individuals have cognitive biases, but that platform design and recommender systems systematically exploit these tendencies at scale. This exploitation emerges from business models that maximise engagement, rewarding emotionally salient and identity-confirming content regardless of its accuracy or democratic value. Even passive consumption of political content may intensify confidence in opinions, making users resistance to compromise and thereby undermining democratic discourse.

Harms to Information Environments

GenAI and social media platforms are transforming the information environment itself. GenAI outputs can promote certain socio-political narratives and produce biased interpretations of historical or political events. Even when these systems appear to report facts neutrally, their outputs depend on training data sets, system prompts, and other choices about which sources are used, how those sources are distilled and recontextualised, and how they are merged with other information. This leads to

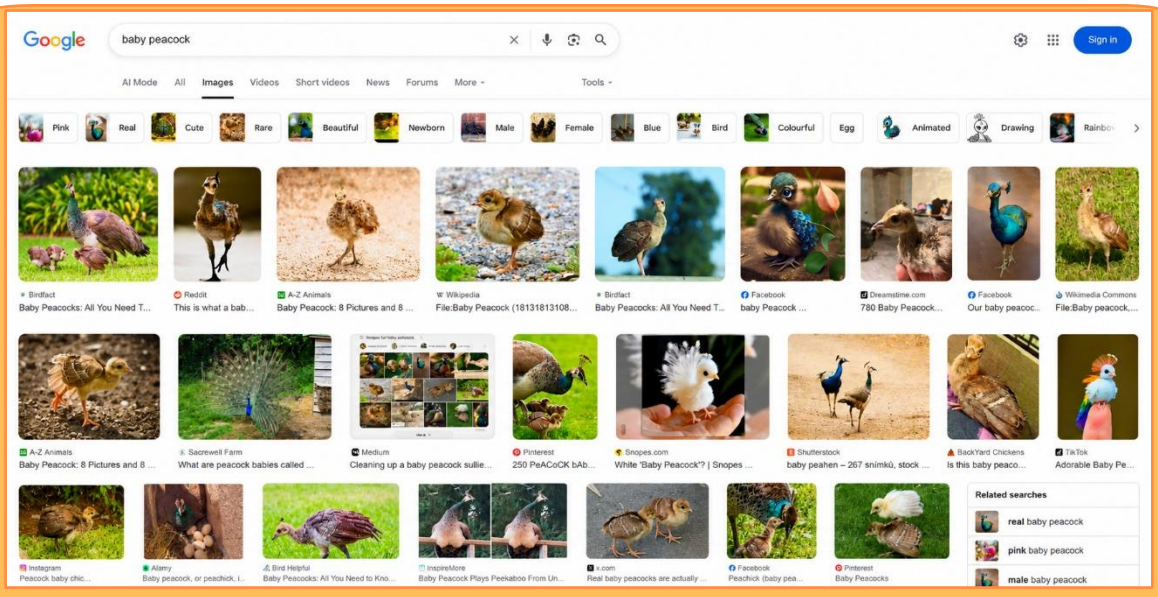

Figure 1. Google Images search results for the query “baby peacock”. The results contain a mixture of photographic and synthetic imagery, illustrating how AI-generated content can become embedded in everyday information environments and complicate users’ ability to assess the accuracy and credibility of visual information. (Screenshot retrieved 29 May 2026)

supposedly neutral outputs that progress specific arguments while obscuring the complexity of a debate while giving the appearance of consensus. Users unaware of how genAI operates often treat the output as neutral and factual.

Recommender systems on social media platforms that determine how content is shared and amplified remain largely opaque. The DSA addresses some of the resulting harms through disclosure obligations about the main parameters of their recommender systems in their terms and conditions (Art 27), requirements that advertisement is identifiable as such and accompanied by basic information about why users are seeing it (Art 26), restrictions on profiling-based advertising to minors (Art 28), and at least one non-profiling recommender options for users of very large online platforms (Art 38).

Despite these provisions, significant practical barriers remain. Article 37 requires very large online platforms (VLOPs) and very large online search engines (VLOSEs) to undergo annual independent audits, while Article 40 enables vetted independent researchers to access platform data. However, recent research on the first round of DSA audits identifies several shortcomings: audit methods differ across platforms, technical assessments often fail to capture the dynamic behaviour of recommender systems, and point-in-time assessments cannot adequately track how these systems evolve between audit cycles. Article 40 provides a second transparency mechanism by allowing vetted independent researchers to access platform data. Yet the accreditation process is lengthy and unreliable, and some platforms have delayed, restricted, or blocked meaningful research access. Article 40 thus provides an important formal right to scrutiny that remains difficult to exercise in practice.

These provisions do not yet provide sufficient oversight of the recommendation logic that determines what content reaches whom and with what cumulative effect. Similar concerns apply to genAI providers, where stronger disclosure of training data and system instructions may be required under the AI Act.

Textbox 2.
How do social media platforms undermine cognitive agency?

Engagement-driven systems may shape belief formation by reinforcing repeated exposure, salience, emotional engagement, and confirmation bias, allowing these mechanisms to function as unreliable proxies for truth. In doing so, they interfere with the intellectual capacities that cognitive agency controls and coordinates, including critical reflection and independent judgement. Overall, this can weaken practices of criticism and engagement oriented toward finding common ground.

Three Policy Recommendations for EU Policy to Safeguard an Informed Citizenry

The AI Act is oriented toward safety, health, and fundamental rights, while the DSA and the Democracy Shield primarily focus on specific instances of illegal content, mis- and disinformation. Neither framework adequately addresses the pervasive transformation of online information environments. GenAI systems and social media platforms need to be regulated as information environments that shape the conditions of democratic citizenship. We propose three specific provisions building on existing instruments that directly protect democracy by safeguarding cognitive agency. Each measure is an important step, but none is sufficient on its own. In the current phase of regulatory simplification, incremental improvements to existing instruments may be more feasible than entirely new regulatory frameworks.

1. Expand systemic risk and manipulation provisions to address harms from synthetic content

GenAI is already regulated under the AI Act as general-purpose AI (GPAI). Applicable obligations include, among others, technical documentation, copyright policies, and, for models posing systemic risks, risk assessments, mitigation measures, and incident reporting to the AI Office. However, current understandings of systemic risk do not adequately capture the harms discussed above. Existing provisions focus on computing power and on harms to identifiable individuals or groups. Mitigation of harms, according to the GPAI Code of Practice, concentrates on reducing false or misleading outputs. This largely leaves unaddressed the large-scale production of synthetic content and the erosion of collective knowledge bases. **EU policymakers should recognise the degradation of information environments caused by synthetic content as a systemic risk in its own right**, so that risk assessments for GPAI include the cumulative effects of synthetic content on collective knowledge bases and information environments.

GenAI also creates new opportunities for manipulation. Highly personalised and persuasive content can be tailored to individual users, making manipulative influences difficult to identify. The AI Act addresses some manipulative practices, but its focus on subliminal techniques and purposeful manipulation leaves important gaps. Manipulation can arise from the routine design and deployment of systems that optimise engagement or reinforce existing beliefs, even where no single output is independently identifiable as manipulative.

EU policymakers should therefore broaden the definition of manipulation to include personalised and cumulative forms of influence that systematically disregard

norms of truthfulness, transparency, and due care. Policymakers should therefore recognise manipulation not only through its immediate effects and through the processes it operates, such as bypassing or undermining the manipulatee's reasoning, but also as a careless influence: goal-directed influence that is indifferent to revealing reasons to its manipulatee and that disregards these crucial norms.

2. Strengthen cognitive agency through educational programs in the Democracy Shield Initiative

The measures discussed above protect cognitive agency indirectly, by improving the environments in which beliefs are formed. Policy should also support citizens more directly. This requires more than general media literacy. Citizens need critical AI literacy so they can understand how AI-generated content is produced, personalised, and presented as authoritative, including the probabilistic nature of genAI and the ways its outputs can reflect socio-political assumptions while appearing neutral. They also need epistemic literacy: the ability to evaluate whether beliefs are justified by appropriate evidence and reasoning, distinguish reliable from unreliable sources, and identify cognitive biases such as confirmation bias and in-group/out-group thinking.

The European Democracy Shield should make the protection of cognitive agency a democratic priority towards strengthening information integrity in Europe. It should pair its focus on disinformation with measures that strengthen citizens' ability to navigate increasingly synthetic information environments. **The EU Digital Education Action Plan 2021-2027 should be expanded to include educational materials on critical AI literacy and epistemic literacy.** Critical AI literacy helps citizens understand how generative AI functions, including its probabilistic nature, technical limitations, and the socio-technical processes shaping its outputs. Epistemic literacy helps citizens evaluate evidence, assess source credibility, recognise when confidence exceeds available evidence, and identify common cognitive biases.

These measures will help citizens exercise better judgment when engaging with AI-generated content and strengthen the cognitive agency upon which democratic participation depends.

3. Improve DSA enforcement to protect information environments through better research access and resources

As noted above, mandatory audits under Article 37 of the DSA have so far proven to fall short in capturing the dynamic behaviour of recommender systems and tracking how these systems evolve between audit cycles. To tackle this problem in practice,

researcher access under Article 40 of the DSA should be simplified and backed by stronger enforcement mechanisms, so that platforms cannot delay or obstruct lawful scrutiny. **The process for recognising vetted researchers should be simplified**, including accepting participation in EU-funded research projects as evidence of research organisation status. **Platforms should also be subject to binding timelines and proportionate penalties for non-cooperation.**

Recent empirical work on TikTok shows why researcher access under Article 40 matters: even when platforms formally comply with the DSA's prohibition on profiling-based advertising to minors, minors may still receive highly personalised commercial content that falls outside the legal definition of advertising. Researchers discovered this DSA infringement through 'sockpuppeting algorithmic auditing' (see **Textbox 3**). By simulating user behaviour through constructed profiles, these methods can reveal behavioural and temporal patterns that formal audits often cannot detect. Here we can gain real insight into which content they show to which users, how often, and under what conditions. Running such research reliably over time requires robust technical infrastructure and sufficient methodological abstraction to generalise across platforms.

Textbox 3. Model-based algorithmic audits ("Sockpuppeting")

Model-based audits test recommender systems by creating controlled user bot profiles, the sock-puppets, and observing what content the platform recommends to them. Researchers use these bot profiles to simulate specific behaviours or interests, then analyse the content shown in response. This helps reveal how recommender systems behave in practice, including patterns that may not appear in formal compliance reports. The method has limitations. Research accounts can be blocked by CAPTCHAs or bot-detection systems; using this method for audits requires substantial technical capacity and expertise; and results can quickly become outdated because platforms and recommender systems change over time.

The EU should fund shared technical infrastructure, secure testing environments, and cross-platform research tools to make model-based auditing, such as sockpuppeting audits, more reliable, scalable, and comparable. **A common set of auditing standards should also be adopted** to improve methodological consistency.

The **three measures proposed here** — expanding systemic risk and manipulation provisions, strengthening cognitive agency through the Democracy Shield, and improving researcher access and independent auditing — are **complementary** and can be implemented **within existing frameworks**. A companion brief will address how the forthcoming Digital Fairness Act can be designed to protect democratic agency in AI-mediated environments.

Center for Environmental and Technology Ethics-Prague (CETE-P)
Institute of Philosophy CAS

Celetná 988/38
Prague 1
Czech Republic

Author Contributions

Conceptualization: Tuğba Yoldaş (online manipulation), John Dorsch (cognitive agency, AI literacy), Jacqueline Bellon (AI literacy, information environment), Jakub Šimko (algorithmic auditing), Matúš Mesarčík (EU policy), Sára Soľárová (EU policy), Andrew McIntyre (information environment). **Funding acquisition**: John Dorsch, Paula Gürtler. **Project administration**: John Dorsch, Paula Gürtler, Tuğba Yoldaş. **Supervision**: John Dorsch, Paula Gürtler, Tuğba Yoldaş. **Visualization**: Anna Kotková, John Dorsch, Paula Gürtler. **Writing – original draft**: Paula Gürtler. **Writing – review & editing**: All authors.

Paula Gürtler, Tuğba Yoldaş, Jacqueline Bellon, Anna Kotková, Andrew McIntyre, Matúš Mesarčík, Jakub Šimko, Sára Soľárová, and John Dorsch. (2026). *Improving Existing EU Policy to Preserve Cognitive Agency: Democracy in AI-Mediated Environments*. Center for Environmental and Technology Ethics — Prague, Institute of Philosophy, Czech Academy of Sciences. https://doi.org/10.5281/zenodo.21278799

AI Disclosure Statement
A local version of EuroLLM-9B was used for minor language editing of the English manuscript and to produce the initial Czech translation. Both versions were subsequently reviewed and revised by the authors, who assume full responsibility for their content. The cover image was created with assistance from Lummi AI and refined in accordance with Lummi's licensing terms.

Funded by the European Union

This project receives funding from the Horizon EU Framework Programme under Grant Agreement No. 101086898. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Center for Environmental and Technology Ethics · Prague

Filosofický ústav AV ČR Institute of Philosophy

Supported by Friedrich-Ebert-Stiftung.

Friedrich Ebert Stiftung