Regulatory Innovations and Policy Options for Synthetic Media and Digital Democracy

Andrew McIntyre,[1] Yasaman Yousefi,[2, 3]
Maria Dolores Sánchez Galera[4]

[1] University of Amsterdam, Amsterdam, the Netherlands
a.mcintyre@uva.nl
[2] DEXAI-Artificial Ethics, Rome, Italy
yasaman.youefi@dexai.eu
[3] University of Bologna, CIRSFID ALMA AI, Faculty of Legal Studies, Bologna, Italy
y.yousefi@unibo.it
[4] Charles III University of Madrid, Madrid, Spain
mariadsa@inst.uc3m.es

This chapter explores potential regulatory innovations and policy options for addressing the democratic risks and opportunities of AI-generated content (AIGC) within the European context. Drawing upon and responding to discussions in previous chapters, it argues that current policy approaches centred on the detection, moderation and containment of AIGC are not only insufficient but also risk reinforcing authoritarian tendencies. Instead, the chapter outlines a policy strategy that emphasizes political participation and pluralism as a means of promoting democratic resilience and addressing the specific harms of AIGC. This strategy is oriented around three key objectives: (i) clarifying AIGC harms, (ii) strengthening institutional coordination, and (iii) enhancing digital literacy and citizenship. Key to this strategy is the reconceptualization of generative AI as a creative and expressive tool for promoting more inclusive political dialogue and democratic debate. Ultimately, this chapter envisions a future in which GenAI is not solely understood as a threat to democracy but as a resource for fostering a more trustworthy information environment and political system. It is a future where truth may become increasingly difficult to determine, but in which our democratic values nonetheless remain protected and strengthened.

DOI
https://doi.org/10.18690/um.feri.2.2026.8

ISBN
978-961-299-109-8

Keywords:
regulatory innovation, counter-disinformation policy recommendation, pluralism, digital democracy, political participation

University of Maribor Press

1 Policy and pluralism

Building on the analysis of democratic risks in Chapter 5 and critiques of mitigation strategies in Chapter 6, this final chapter examines how harmful AI-generated content (AIGC) is conceptualised in current European policy and proposes new governance strategies. To begin, section 8.1 explores the unique challenges of counter-disinformation policy, showing how measures aimed at governing truth may erode democratic trust and promote authoritarian tendencies, highlighting the need for active citizenry and pluralist debate. Beyond addressing the negative impacts of AIGC, section 8.2 then considers how GenAI could be utilised as a unique tool of representation and communication that can promote pluralist debate and political participation. Finally, section 8.3 builds on these discussions to outline priority areas for policy as part of a broader strategy that addresses harms while promoting democratic resilience. This requires clarifying harms, acknowledging tensions, and reconceptualising AIGC as socio-political resources rather than solely risks that need to be mitigated.

Before discussing European policy specifically, it is necessary to briefly frame this policy discussion within the broader epistemic context of GenAI. As Floridi argues, we now exist in an infosphere where human experience and knowledge are redefined in terms of information flows (Floridi, 2014). From this perspective, AIGC does not simply mislead individuals; it contributes to and alters the structural integrity of our wider information environment (Russo, 2022). Beyond introducing artificial content, AIGC reshapes the epistemic conditions under which societies construct, verify, and contest knowledge (Bisconti et al., 2024). Disruption has profound implications for collective knowledge, socio-political discourse, and democratic deliberation (McIntyre et al., 2025). AIGC is not inherently detrimental, but its use for disinformation presents what we describe as *informational harms*.

As Feinberg argues, harm is a wrongful infringement or obstruction of a person's interests. These interests include one's physical safety and further extend to other interests such as property, privacy, autonomy, and reputation, among others. Therefore, harm can be both tangible (e.g., physical violence, theft) and intangible (e.g., violating privacy, restricting autonomy) (Feinberg, 1987). Within Floridi's infosphere, however, human beings are redefined as informational organisms whose identity, agency, and interests are fundamentally constituted by information flows

and structures within our broader informational environment. Through this theoretical lens, we reconceptualise Feinberg's notion of harm as an infringement or obstruction of a person's informational integrity. As a person's informational being is embedded within and continually shaped by the wider infosphere, however, protecting individuals from harm ultimately depends on maintaining the integrity of the information environment as a whole. Thus, informational harms relate to how people are impacted by deception, misrepresentation, and disinformation, and how processes of knowledge construction, dissemination, and reception are impacted by the social integration of AI systems and the widespread production of AIGC.

To translate the notion of informational harms into policy, we draw on Smuha's harm categories related to AI. As Smuha argues, harms can be categorised at three levels: (i) individual, when people are directly misled (e.g., deceptive deepfakes); (ii) collective, when groups are disproportionately affected (e.g., racial stereotypes); and (iii) societal, when institutions and governance are undermined (e.g., synthetic media in elections) (Smuha, 2021). For example, the 2024 US presidential election, marked by a surge in AIGC, exemplifies societal harms by eroding trust in institutions. The EU recognizes such risks in the AI Act, which acknowledges GenAI may generate material or immaterial harm (European Union, 2024). Yet existing frameworks remain reactive, focusing on moderation and detection rather than systemic impacts.

This chapter outlines policy priorities that address harms across these different levels while grappling with tensions such as institutional dysfunction and reconciling regulation with freedom of expression. Confronting these directly, the chapter offers a blueprint for reconceptualising AIGC as a potential resource for democratic resilience.

The European legal mechanisms discussed in Chapter 7 offer only limited solutions to the significant challenges posed by harmful AIGC. Many of these mechanisms are narrow in scope and practical application, failing to fully account for the deep integration and diverse use of GenAI in everyday life. As such, these frameworks do not adequately define or conceptualise AIGC as a socio-political phenomenon, nor do they address the diverse harms that AIGC can inflict upon different levels of society (individual, collective, societal). In section 8.3, we elaborate on possible legal innovations to more appropriately address the harms associated with AIGC as part of our wider policy priorities. However, legal solutions alone cannot fully account

for the deep social integration and diverse use of GenAI in everyday life. As such, we need more diverse policy interventions and strategies for combating the spread and impact of harmful AIGC, as well as solutions for promoting stronger democracies.

Broadly speaking, emerging policy strategies fall into one of three categories: (i) retreat strategies aimed at reducing digital interactions in favour of in-person interactions to improve trust relationships; (ii) containment strategies aimed at detecting, labelling and limiting the impact of harmful AIGC; and (iii) mobilization strategies aimed at harnessing GenAI to promote more robust democratic systems (Allen & Weyl, 2024). Largely, states have pursued containment strategies as they focus on practical and tangible technological, legal, and social solutions and allow for the strict regulation of harmful AIGC. However, though well-intentioned in their attempt to protect informational integrity and democratic stability, many of these containment strategies seek to re-establish a single authoritative source of truth and, in doing so, paradoxically undermine democracy while reinforcing anti-democratic tendencies. To elaborate, let us critically examine the goals and assumptions underpinning these strategies, which Farkas and Schou divide into four dimensions: (i) policing the truth; (ii) re-establishing centres of truth-making; (iii) promoting public immunity; and (iv) technological solutionism (Farkas & Schou, 2023).

To elaborate, many containment strategies are aimed at policing truth, often relying on restrictive legislation and other drastic measures that policymakers justify as protecting the democratic foundations of truth and reason. However, Farkas and Schou describe such measures as authoritarian in that they are veiled attempts at censorship that consolidate government control over the information environment. Furthermore, these strategies shift open political debate into closed governmental mechanisms, which are rarely subject to public scrutiny. Secondly, often these efforts aim to re-establish traditional centres of truth-making (e.g., politics, science, journalism) and position these institutions as vital protectors of truth that must reclaim authority. Science, in particular, is often privileged above others, with researchers and technologists arguing that they should be included in high-level decision-making, even to the point of superseding public opinion. However, Farkas and Schou claim that these approaches risk emboldening certain groups as arbiters of truth, reinforcing the elitist notion that governance should be dictated by technocratic experts rather than public dialogue. Similarly, public education

initiatives (e.g., media literacy programmes) aimed at strengthening individual critical thinking are certainly important and beneficial. However, these strategies are often framed as a method of curing public ignorance or immunising the public against manipulation. Farkas and Schou argue that such a framing places responsibility on individuals rather than governments or technology companies, while also dismissing popular dissent and diverse opinions as ignorance or delusion that is simply wrong in comparison to the single truth defined by experts.

Such strategies also often utilise advanced technologies, including AI systems, in order to detect, verify, and manage disinformation. While certainly technical innovations can be effective and beneficial, often these technical fixes are presented as the only viable solution and are too simplistic to fully address nuanced socio-political challenges. Furthermore, this relies upon private technology companies and gives these companies control over what constitutes truth and societal harm (Allen & Weyl, 2024).

This is not to say that technological solutions are inherently problematic and, indeed, we advocate for the ethical and transparent use of AI systems below. However, we wish to highlight that the blunt use of technologies to determine truth and harm risks undermining democracy further.

While we largely agree with Farkas and Schou's critiques and agree that we should not be attempting to arbitrate truth, we would not fully condemn or abandon these containment strategies.

These strategies offer partial solutions, but in the rush to combat disinformation, they may inadvertently undermine the very democratic values they seek to protect. The challenge, therefore, is not to discard these policies altogether but, rather, to implement them with a heightened awareness of the risks and ensure that they are designed to promote a more resilient, rather than a more controlled democracy.

This approach forms the core of the policy priorities presented in section 8.3 of this chapter. However, we must go further than simply careful and ethical implementation of containment strategies that seek to determine and arbitrate truth. As Farkas and Schou argue, we require an alternative approach for strengthening democracy that is not about establishing a single truth at all. Instead, they advocate

for a pluralistic and genuinely political public sphere that embraces the "always-antagonistic dimension of the political" by fostering "spaces for vibrant clashes of conflicting alternatives" (Farkas & Schou, 2023).

Drawing on the work of political philosophers like Chantal Mouffe (Mouffe, 1997), Ernesto Laclau (Laclau, 1990), and Jacques Rancière (Rancière, 2014), Farkas and Schou contend that the current post-truth political crisis is not due to a lack of facts or an increase in deceptive media. Instead, it stems from a lack of meaningful democratic participation. More specifically, they argue that a healthy democracy is not about reaching a rational consensus on what is true but, rather, about embracing a culture of constructive and agonistic pluralism that involves a vibrant clash of democratic political positions. Therefore, instead of focusing solely on counter-disinformation measures, Farkas and Schou argue that policymakers should couple these measures with strategies that encourage greater and more diverse political participation; "more politics" rather than "more truth".

With the arrival of GenAI, we are fast approaching a world in which everyday people, not only states and companies, are powerful media producers capable of creating and distributing convincing AIGC around the world in moments. In such a world, retreat strategies are impractical and potentially detrimental in that technology bans are unlikely to be adopted by states, and it is unrealistic to expect people to voluntarily abandon digital life.

Even if this were achieved, they risk undermining the positive political uses of digital technologies (e.g., increased communication and representation), while squandering further potential uses of GenAI. Furthermore, containment strategies can only go so far and risk fostering authoritarian tendencies and exacerbating distrust in democratic institutions, as discussed. If we accept that the proliferation and social integration of GenAI will continue at pace, we cannot solely rely on retreat or containment strategies. Instead, it is necessary to embrace mobilization strategies that utilise GenAI to promote political engagement and agonistic pluralism. Where Allen and Weyl highlight the use of such systems for authentication, data privacy, and promoting access to public information spaces, we contend that AIGC can play a role in this constructive agonistic dialogue and could be used to promote democratic resilience.

2 Deepfakes for political participation

Much attention has been paid to the negative impacts of AIGC, and rightly so, given their origins in deepfake pornography and the imminent threats they pose to democracy. Not only does AIGC risk misrepresenting the actions and statements of individuals, but it also impacts the integrity of our information environment and disrupts communication between citizens or groups of citizens, thus undermining democratic processes of collective decision-making. As Mathias Risse argues, for citizens to make collective decisions on policies and laws that will affect the population, they require "a decent level of knowledge about the people with whom they share a polity, lest these citizens be deceived, e.g., about how certain measures affect others or what such people's worries are (Risse, 2023). Harmful or deceptive AIGC may lead to greater misunderstandings or animosity between different communities, encouraging political polarization that stifles collaboration and dialogue. However, there are more diverse uses of AIGC that have received less public attention but that indicate how GenAI could be utilised to promote democratic values and political participation.

This discussion focuses on those instances in which AI-generated content has been used to improve public engagement with socio-political discourse and/or encourage communication and empathetic connection between citizens. These instances might include, for example, translating government communications to engage with multi-lingual communities (e.g., Manoj Tiwari speaking Haryanvi in 2020 (Jee, 2020)), creating interactive education tools or exhibitions to better explain historical events and figures (e.g., *Dalí Lives* exhibition (Lee, 2019)), or visualising future scenarios to better communicate the consequences of abstract policy issues (e.g., *This Climate Does Not Exist* (Tousignant, 2021).

A particularly illustrative example is the exhibition *EXHIBIT A-i* (Blackburn 2023), which used GenAI to visualise the witness statements of 32 refugees previously held at Australia's offshore detention centres on Manus Island and Nauru (Doherty, 2023). Gathered by the law firm Maurice Blackburn, these witness statements explained in graphic detail the inhumane conditions of these centres and the regular incidents of violence, abuse, self-mutilation, rape, and suicide that occurred there. As reporters were restricted from accessing these centres, no photographs or recordings exist, and so a text-to-image GenAI system was used to produce visual

representations. It is important to note that these synthetic images were not intended as deception or as a substitute for evidence and their artificiality is openly acknowledged in the exhibition. Regardless, these artificial images provide the public with a bleak and visceral depiction of life in these centres and thus enable a more intimate understanding of the experiences of real people than can be achieved through text alone. Such images emphasize the human and personal impact of immigration policies, thus allowing citizens to better assess the actions of government institutions and the choices made by those politicians and officials in positions of power.

While these positive uses of AIGC are currently rare and often regarded as little more than curiosities or artistic experiments, they highlight the potential of how GenAI might be used to improve socio-political participation and epistemic agency. With greater and more engaging access to information about historical events, other communities, and the real and potential impacts of said policies on different communities, citizens may be able to more effectively formulate their own political opinions, empowering them to more competently engage with political discussions and to more confidently exercise their political agency in collective decision-making processes.

In Chapter 6, we explored the use of AI-generated content to promote specific values that aligned with the United Nations Sustainable Development Goals (SDGs). While they offer a creative and engaging way of communicating the SDGs, many participants in our use case expressed concern about the potential for deception and political manipulation, as well as the ethics of using historical or deceased figures to promote certain ideas without consent. These concerns echo those of Farkas and Schou with regard to authoritarian tendencies and the policing of truth. Rather than utilise GenAI to communicate selected values perceived as democratic (e.g., SDGs), it seems more appropriate and more democratic to place these technologies in the hands of citizens themselves and to encourage ethical use in public communication. As this technology becomes more deeply embedded into our everyday lives and communicative practices it has the potential to strengthen pluralist debate and remove barriers to political participation.

Previously, a lack of resources (e.g., finances, time, technology) or limited communicative capabilities (e.g., storytelling, oratory, technical skills) might have restricted citizens from fully participating in democratic dialogue and decision-making. With GenAI more widely available, however, the average citizen needs only provide a simple prompt to rapidly produce expressive, empathetic, and engaging audiovisual content representing their daily life. In doing so, individuals could easily visualise their personal experiences and private events that might otherwise go undocumented or ignored. This could include instances of systemic violence, abuse, and neglect, ensuring that the injustices and inequalities that citizens endure are visualised in detail, in ways that resonate with the wider public.

This is not to argue for a purely technological solution but rather to highlight how such technologies might be utilised through mobilization strategies to promote democratic values. Certainly, the widespread use of GenAI has significant risks (e.g., pornographic abuse, disinformation), but if appropriately implemented, this technology could enable citizens to better appreciate the lives of other communities, to engage with a plurality of views, and to understand how government policies and legislation might impact one another differently. Recalling Farkas and Schou's constructive antagonism, the purpose of such strategies is not to arbitrate truth but, rather, to promote a more vibrant, creative, and plural political debate. Coupled with light-touch containment strategies and legislative innovations, we may begin to move toward a more trustworthy information environment and political system wherein truth may become increasingly difficult to ascertain but wherein our democratic values are nonetheless upheld. The use of AIGC for promoting political engagement, alongside containment and literacy strategies, forms a key aspect of our proposed policy priorities described in the next section.

3 Regulatory and policy priorities for democratic resilience

Based on the above discussion, we propose that a strategy for democratic resilience should be aimed at maintaining the integrity of our information environment and, rather than arbitrating the truth, promoting a technically literate and politically active citizenry. While we recognise the need for containment strategies and technological solutions, this strategy emphasizes societal adaptation through conceptual unity in law and policy, robust democratic systems, and social integration of AI. This strategy builds upon the specific measures recommended by the European Parliamentary

Research Service (EPRS), as well as other existing counter-disinformation policy and regulatory proposals. It aims to address harms across Smuha's three levels of harm (individual, collective, societal) and is oriented around three key objectives: (i) legal clarification of AIGC and informational harms; (ii) coordination of democratic institutions; and (iii) promoting plural and participatory citizenship. These priority proposals are explained in more detail below, while Table 8.1 illustrates how they are aligned with the strategic objectives and how they address the levels of harm.

Table 1: Priority proposals for democratic resilience

Objectives	Priority Proposals	Harm level		
		Individual	Collective	Societal
Clarification	Unified legal framework	(x)	(x)	(x)
	Unified personality rights	(x)		
	Transparency obligations	(x)		
Coordination	Unified infrastructural investment			(x)
	Multi-stakeholder coordination			(x)
Citizenship	Media and AI literacy	(x)	(x)	
	Technical citizenship	(x)	(x)	(x)
	Pluralist media landscape		(x)	(x)

Source: Own

3.1 Unified Legal Framework on Synthetic Media

Across European legislation, policy, and counter-disinformation strategies, the specific issue of AIGC is ill-defined. In the context of AI governance legislation and policy (e.g., AI Act, national AI strategies), the harms of AIGC are noted as a concern, but other socio-political issues (e.g., algorithmic bias, surveillance) are often prioritized. Meanwhile, counter-disinformation strategies often equate AIGC with traditional forms of disinformation, and it is often assumed that current tactics can be simply extended such that there are little to no explicit policies or strategies aimed directly at AIGC as a distinct problem requiring specific responses, as many experts have called for.

This ambiguity around the issue of disinformation further extends to how the problem is conceptualized more broadly. In terms of scale, disinformation can be understood as a problem in which harmful individual content spreads naturally

between online users and thus requires more robust moderation mechanisms; such is the approach of the UK Online Safety Act. However, the national strategies of countries such as Spain (Gobierno de España, (2019) and France (Ajji, 2020) conceptualise disinformation as a coordinated and motivated campaign involving the spread of harmful narratives through numerous pieces of online content and thus require a national response. Furthermore, many of these strategies focus on the issue of electoral interference while overlooking the continual role that disinformation plays in everyday abuse, encouraging polarization between communities, and eroding confidence in democratic institutions.

With these different conceptualizations of disinformation comes further ambiguity around what constitutes harmful content. Notably, the UK Online Safety Act identifies harmful content as that which causes psychological or physical harm upon an individual, while the Digital Services Act (DSA) considers the broader societal harms of disinformation and other national criminal codes, such as those in Italy, Spain, and Albania, characterise harm in terms of public order and citizen safety.

Most critically, counter-disinformation policy must navigate the fundamental tension with freedom of speech. The boundary between harmful disinformation and protected speech is often blurred, and any policy, even one that is non-legislative, runs the risk of creating a chilling effect on legitimate expression. As discussed, a focus on banning or removing content can lead to further public distrust in regulatory institutions and can be easily co-opted by authoritarian regimes to suppress dissent.

Given these complexities and ambiguities, existing laws addressing harmful online content must be updated to address the specific challenges of harmful AIGC, and particularly, they require a clearer definition of what constitutes disinformation and what constitutes harm. We propose establishing a taxonomy of disinformation based on the semiotic models discussed in Chapter 3 and clearly identifying AIGC within this taxonomy. Such a taxonomy differentiates disinformation that is based on falsification of the material form (e.g., manipulation or fabrication) and that which is based on falsification of the content (e.g., misrepresenting authentic content). Harmful AIGC falls into the first category. Based on these categories, more specific definitions and guidelines can be established.

As the EPRS recommends, clearer guidelines are necessary for applying the General Data Protection Regulation (GDPR) framework to deepfakes, while strengthening the capacity of data protection authorities to address unlawful data processing, and developing a unified approach to personality rights within the EU (discussed below). Furthermore, we should protect the personal data of deceased persons, for example, with a "data codicil" and institutional support for victims of AIGC by providing accessible judicial and psychological resources.

Given the role that AIGC plays in individual harms (e.g., pornographic abuse), collective harms (e.g., political polarization), and societal harms (e.g., distrust in institutions), a unified strategy is crucial to addressing all three levels.

3.2 Unified Personality Rights

Similarly to definitions of AIGC and harms, personality rights covering an individual's name, likeness, image and voice are currently not harmonized at the EU level. This leaves regulation to the discretion of Member States and resulting in a patchwork of approaches. For example, France protects personality rights primarily through privacy and image rights, while Germany provides stronger safeguards by recognising personality rights under its constitution. By contrast, the UK lacks standalone legislation to cover personality rights but, instead, relies on a combination of privacy law, defamation, and tort law.

As the harms of AIGC transgress national boundaries, the EU should harmonize regulations related to personality rights to ensure consistent protection of citizens and to prevent malicious actors from exploiting these regulatory differences. A potential grounding for EU-level personality rights could be the recently proposed amendment to the Danish Copyright Act that is explicitly designed to address the issue of AIGC and digital imitations (Denmark, 2023).

This draft law treats identity as intellectual property and aims to give citizens copyright-style rights over their own likeness, voice, and physical features. Under the proposal, citizens can demand the removal of AIGC, representing themselves, made without consent, and seek compensation, even if no reputational damage is proven. Online platforms would be legally required to take down such content once notified or face sanctions, while carve-outs remain for free expression uses such as

parody and satire. The law also offers specific protection to performing artists against unauthorized digital reproductions of their work. Broadly, this approach could be expanded across the EU to give citizens an explicit legal mechanism for controlling their own likeness and for combating individual harms of AIGC.

This could be achieved by updating existing legislation. Firstly, the EU Copyright Directive should be updated to give citizens the right to their own likeness, similarly to performers. Secondly, the GDPR should be updated to redefine AIGC that replicates an individual's likeness or voice as protected personal data, even if created entirely synthetically. Finally, the AI Act's transparency obligations could be expanded to include individual consent and rapid takedown rights. Together these updates would create robust regulation for preventing misrepresentation through AIGC.

3.3 Transparency Obligations

While the AI Act introduces transparency obligations to clearly label deepfakes circulating on online platforms, further transparency obligations should apply to AI moderation and deepfake detection systems used by these platforms. As discussed in 8.1, these technological containment measures risk being perceived by the public as authoritarian attempts at censorship that police the truth and insist upon a single arbiter. Without transparency, the use of AI systems to restrict the spread of harmful content may backfire causing further public distrust of governments and organizations. To combat this, we propose that platforms be required to disclose how their AI moderation and deepfake detection systems operate. This transparency would allow users to understand how content is moderated and flagged, while also providing a basis for holding platforms accountable for their decisions. Clear procedures for labelling deepfakes and a robust appeal mechanism must be established to ensure fair treatment and protect legitimate uses of GenAI.

3.4 Unified Infrastructural Investment

All of these strategies depend on strong government and private organizations, nationwide organizational networks, substantial funding, and the technical infrastructure needed for implementation. While robust policy frameworks may succeed in developed nations with sufficient capacity, they are often unworkable in

regions with low digital literacy, limited access to technology, and weaker government systems. This digital divide is a major barrier to a unified European approach, and a major challenge to the integrity of our broader information environment and leaves us all more vulnerable to harmful AIGC. We must address this divide through international cooperation and investment programmes that build foundational digital infrastructure and establish comprehensive regulatory systems.

Without such efforts, proposed solutions risk deepening existing inequalities and failing to address the global scope of the threat. The EPRS (van Huijstee et al., 2021) highlights one response: authentication systems that enable users to verify content through digital watermarks or registered information provenance, extending also to court evidence. It further recommends coordinated investment in AI systems for detection and prevention, alongside diplomatic measures and international agreements to deter foreign state actors, reinforced where necessary by economic sanctions. To close capacity gaps in organizations and developing nations, the EPRS also calls for investment in knowledge and technology transfer, and for both public and private entities to conduct their own risk assessments. Primarily, this measure addresses broader societal harms of deepfakes and synthetic media by seeking to give all Member States and institutions sufficient tools to tackle disinformation across borders.

3.5 Multi-stakeholder Coordination

As discussed in Chapter 2, harmful AIGC can rapidly spread throughout online networks, and so it is necessary to establish early-warning systems that integrate technical and human intelligence. A primary obstacle to effective counter-disinformation strategies is institutional dysfunction (e.g., different standards and definitions for disinformation) and a lack of collaboration between key stakeholders across society, such as platforms, governments, research institutions, and media organizations. For example, governments may be hesitant to share sensitive data with private companies, while platforms may be unwilling to share proprietary data with public research institutions. Policy can attempt to bridge these gaps by establishing neutral, third-party convenors and by creating a clear set of shared ethical principles that all parties agree to uphold. This lack of collaboration and coordination is also evident between local, national, and European-level organizations, where differing policies, jurisdictions, and resources create

inefficiencies. Some states have sought to tackle this issue directly. Notably, Spain's Protocol to Combat Disinformation (Gobierno de España, 2021) emphasizes inter-agency cooperation, while the UK has introduced regional cybersecurity hubs to coordinate responses, primarily to cyber threats and to disinformation instances (UK Government, 2022). However, many other states suffer from a lack of coordination. In particular, this dysfunction hinders efforts to rapidly address large-scale infodemic scenarios involving AIGC.

To address this dysfunction, key government institutions, social media platforms, fact-checking groups, and media organizations at the local, national, and international levels should establish a unified counter-disinformation network. Such a network would enable a real-time infodemic alert system whereby harmful AIGC identified by one organization can be immediately flagged for review by all partnering organizations, and the network as a whole can launch simultaneous public awareness campaigns to highlight the infodemic risk to citizens. Such a network approach would foster a transparent, agile verification process that allows multiple perspectives to contribute without resorting to heavy-handed state policing of the truth. Furthermore, the interconnected and multi-level nature of this approach would more effectively tackle infodemic events by enabling rapid verification and widespread public communication. This creates a network effect of protection, where the detection of a single piece of harmful content by one entity contributes to the resilience of the entire ecosystem, thus moving from a fragmented and reactive response to a more proactive and coordinated defence.

Key to this counter-disinformation network is increased investment in local journalism and media organizations that are trusted within their immediate communities. With increased funding, local media could provide reliable firsthand reporting that feeds into national and international levels, while also playing a direct role in public communication and serving as trusted intermediaries between the local community and the wider information ecosystem. Such investment would also be bolstered by greater coordination with online platforms to ensure citizens receive localized news. Furthermore, the use of local media organizations instead of government communication hubs ensures independence and avoids authoritarian tendencies.

While challenging to implement due to institutional dysfunction and lack of resources, this network approach is an effective way to address the multi-level, cross-border, and cross-platform nature of disinformation threats.

3.6 Media and AI literacy

For decades, there has been a strong emphasis on building public resilience to political manipulation through media literacy initiatives at both national and EU levels. Such efforts remain essential to democratic resilience in the age of synthetic media, empowering citizens to be active and critical participants in socio-political discourse.

However, current initiatives often lack a specific focus on GenAI, and so literacy programmes need to evolve to respond to our continually changing information environment. As the EPRS recommends, AI literacy should be integrated into formal educational curricula from a young age in order to teach students how to critically consume synthetic media and how to analyse its production, purpose, and potential harms (van Huijstee et al., 2021).

This includes teaching citizens how to identify AIGC (e.g., unnatural eye movement, distorted backgrounds, audio glitches), as well as a broader understanding of how GenAI systems are trained and the biases they may contain. Moreover, literacy programmes should teach citizens to recognise AI-generated content based on technical and, furthermore, encourage citizens to consider the context, such as the content's source and broader background information about the people and events they are shown.

This does not simply require more general media literacy training, and requires citizens to be more deeply engaged with politics and events. Furthermore, AI literacy initiatives should engage citizens across all stages of life, from primary education to professional training and adult programmes. Meanwhile, targeted programmes should seek to engage vulnerable groups who may lack certain literacy skills, such as older adults or people with learning and cognitive disabilities.

Promoting AI literacy is not only an effective strategy for combating individual manipulation or deception, but, if implemented consistently across society, such initiatives address those broader epistemic and societal harms caused by

disinformation. By equipping citizens with the ability to discern reliable information from synthetic noise, we can begin to rebuild trust in democratic institutions and political processes. While such initiatives should receive government funding, independent educational institutions and citizen science organizations must implement AI literacy programmes to avoid the perception of authoritarian arbitration of truth that Farkas and Schou highlight. Such programmes can lead to an AI-literate citizenry that is more resistant to manipulation. If coupled with technical citizenship initiatives, as the next section will explain, this could further encourage a more vibrant AI-enabled public discourse and political participation.

3.7 Technical citizenship

To encourage a more vibrant and active political participation, AI literacy programmes need to go beyond simply teaching ways of identifying AIGC and critical engagement with GenAI. These programmes should also focus on ethical and pro-democratic use of such technologies that do not focus on deceptive practices but, rather, methods of AI-enabled personal representation and self-expression. Investing in this more practical curriculum is to cultivate a citizenry that is AI literate and aware of the technology's societal impacts, and is also utilising AI positively and actively engaging in plural democratic debate. It is important for these initiatives not to simply encourage greater use of GenAI but to emphasise the ethical use of these technologies for personal representation and self-expression rather than manipulative deception.

Beyond further investment in formal education programs for technical citizenship, policy can be used to promote informal and community-driven initiatives. Policy support could include publicly funded online spaces or channels for teaching AI literacy and ethical use, as well as grants for community-based organizations to host workshops and information sessions, particularly in marginalized communities disproportionately affected by disinformation campaigns (Gautam et al., 2024). Such sessions could focus on creating online spaces wherein citizens can participate in political discussions in creative and empathetic ways by utilising AI-generated content. Platforms such as YouTube and GitHub could also be repurposed as such spaces for public engagement (McCosker, 2024).

Combined with media and AI literacy, technical citizenship initiatives are intended to encourage a more trustworthy information environment and to promote a pluralist media landscape in which citizens are politically engaged and where numerous different socio-political views are represented.

3.8 Pluralist media landscape

Beyond literacy and technical citizenship initiatives, a significant obstacle to implementing counter-disinformation strategies is the increasingly fragmented media landscape across Europe and within individual Member States. The widespread availability of digital technologies and the rapid growth of social media have drastically increased the number of people capable of producing and disseminating information online. As such, many users and entire communities no longer share common sources of information, instead consuming highly personalized content shaped by recommendation algorithms. This explosion of online platforms makes it difficult to monitor information flows and ensure compliance with counter-disinformation legislation. Notably, the provisions of the DSA only apply to very large online platforms, leaving smaller but still influential sources largely unregulated.

Countering disinformation requires a strong, diverse media ecosystem. Policymakers should support independent journalism and media organizations to ensure that the public has access to reliable, high-quality information, while also supporting pluralistic debate. Promoting diverse media sources and critical reporting can help resist the normalization of biased or distorted narratives through AIGC, without resorting to authoritarian overreach.

A key component of this approach is addressing capacity gaps that exist in smaller media organizations and civil society groups that are essential for ensuring diverse perspectives. Policy could establish national or international funds, supported by government grants and philanthropic contributions, to provide these organizations with access to advanced tools and training. This would ensure that the ability to combat disinformation is not a luxury reserved for well-funded entities, but a widely distributed capability that strengthens the entire information ecosystem. Crucially, this approach avoids the centralization of media power, instead fostering a plural and resilient information ecosystem.

4 Concluding Remarks

Any comprehensive strategy that aims to effectively regulate against the harms of AIGC in the European context must first recognise that these harms are rooted in the degradation of our information environment. Accordingly, the harms posed by AIGC are not solely related to misrepresentation or deception of individuals, but rather they relate more broadly to the integrity of collective knowledge and manifest differently across different levels of society (individual, collective, societal).

Existing EU legislation remains fragmented and inadequate when addressing this specific issue, and there is an urgent need for more clarity. However, legal tools alone are insufficient to address the deep social integration of these technologies into our social lives and the diverse harms this integration presents. Additionally, these legalistic approaches do not fully embrace the potential opportunities for using GenAI to revitalise plural political debate. To properly address this issue, policymakers should adopt a holistic approach that balances technical and legal solutions aimed at containing disinformation with pluralist social policies aimed at promoting political participation.

In this chapter, we developed an approach oriented around three primary strategic objectives: (i) clarifying harms of AI-generated content through unified legal definitions and personality rights; (ii) strengthening institutional coordination through multi-stakeholder collaboration and investment; and (iii) enhancing citizenship through AI literacy, technical skills, and a plural media landscape. Rather than viewing AIGC solely as a threat to be contained through heavy-handed measures, regulatory and policy innovations should focus on adapting society around GenAI. Central to future democratic resilience is the cultivation of a technically literate and politically active citizenry that is able to recognise and resist AI-generated disinformation and actively uses GenAI tools to contribute to the political debate.

End notes

Andrew McIntyre conceptualized the Chapter and coordinated the writing. He wrote the introduction and conclusions, deepfakes for political participation, Regulatory and policy priorities for democratic resilience. Yasaman Yousefi wrote the section on Policy and Pluralism. She also contributed to the section on Regulatory and policy priorities for democratic resilience, specifically to the legal analysis

and policy recommendations. Maria Dolores Sánchez Galera contributed to the legal analysis. All authors reviewed and approved the final version.

References

Ajji, K. (2020). Protecting liberal democracy from artificial information: The French proposal. In B. Petkova & T. Ojanen (Eds.), Fundamental rights protection online (pp. 57–83). Edward Elgar Publishing. https://doi.org/10.4337/9781788976688.00013

Allen, D., & Weyl, E. G. (2024). The real dangers of generative AI. *Journal of Democracy, 35*(1), 147–162.

Bisconti, P., McIntyre, A., & Russo, F. (2024). Synthetic socio-technical systems: Poiêsis as meaning making. *Philosophy & Technology, 37*(3), 94. https://doi.org/10.1007/s13347-024-00730-y *(DOI added when available – remove if you prefer strictly source-based)*

Codice Penale (Italia), R.D. 19 ottobre 1930, n. 1398, arts. 656, 595, 612.

Criminal Code of the Republic of Albania, Law No. 7895 of 27 January 1995, as amended by Law No. 146/2020, arts. 267, 271.

Denmark. (2023). Consolidated Act on Copyright (Consolidated Act No. 1093 of August 20, 2023). WIPO Lex. https://www.wipo.int/wipolex/en/legislation/details/22692

Doherty, B. (2023, April 8). Manus Island and Nauru: Previously unseen testimony and AI imagery reveal "unimaginable" part of Australian history. *The Guardian*. https://www.theguardian.com/australia-news/2023/apr/08/manus-island-and-nauru-previously-unseen-testimony-and-ai-imagery-reveal-unimaginable-part-of-australian-history

European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. *Official Journal of the European Union, L 168*, 1–135.

Farkas, J., & Schou, J. (2020). *Post-truth, fake news and democracy: Mapping the politics of falsehood*. Routledge.

Feinberg, J. (1987). *Harm to others*. Oxford University Press.

Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*. Oxford University Press.

Gautam, A., Joshi, R. K., Narula, A., & Sharma, N. (2024). Mitigating human rights violations caused by deepfake technology. *Library Progress (International), 44*(3), 4628–4637.

Gobierno de España. (2019). *National Counter-Terrorism Strategy 2019*.

Gobierno de España. (2021). *Estrategia de Seguridad Nacional 2021*.

Jee, C. (2020, February 19). An Indian politician is using deepfake technology to win new voters. *MIT Technology Review*. https://www.technologyreview.com/2020/02/19/868173/an-indian-politician-is-using-deepfakes-to-try-and-win-voters/

Laclau, E. (1990). *New reflections on the revolution of our time*. Verso.

Lee, D. (2019, May 10). Deepfake Salvador Dalí takes selfies with museum visitors. *The Verge*. https://www.theverge.com/2019/5/10/18540953/salvador-dali-lives-deepfake-museum

Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales. (2018). *Boletín Oficial del Estado*.

Loi n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information. (2018). *Journal officiel de la République française*.

McCosker, A. (2022). Making sense of deepfakes: Socializing AI and building data literacy on GitHub and YouTube. *New Media & Society, 26*(5), 2786–2803.

McIntyre, A., Conover, L., & Russo, F. (2025). A network approach to public trust in generative AI. *Philosophy & Technology. (Advance online publication – update when vol./issue available)*

Mouffe, C. (1993). *The return of the political*. Verso.

Nemitz, P. (2022). People or technology: What drives democracy? *Transatlantic Policy Quarterly, 20*(4), 35–42.
Rancière, J. (2014). *Hatred of democracy*. Verso.
Risse, M. (2023). *Political theory of the digital age: Where artificial intelligence might take us*. Cambridge University Press.
Russo, F. (2022). *Techno-scientific practices: An informational approach*. Rowman & Littlefield.
Smuha, N. A. (2021). Beyond the individual: Governing AI's societal harm. *Internet Policy Review, 10*(3). https://doi.org/10.14763/2021.3.1579
Tousignant, B. (2021, October 13). This climate does not exist: Picturing impacts of the climate crisis with AI, one address at a time. *Mila*. https://mila.quebec/en/article/this-climate-does-not-exist-picturing-impacts-of-the-climate-crisis-with-ai-one-address-at
UK Government. (2022). *Government Cyber Security Strategy: Building a cyber resilient public sector*.
UK Government. (2023). *Online Safety Act*.
Van Huijstee, M., van Boheemen, P., & Das, D., et al. (2021). *Tackling deepfakes in European policy*. European Parliamentary Research Service.