Report III DeTS

From Discourse to Governance - AI, Language, and Democratic Resilience.

Final Findings and Policy Guidance (DeTS): Three Pillars for Democratic Resilience in the AI Public Arena

28/02/2026

By Andrea Declich and Fabio Feudo, in cooperation with Monika Banaś

Overview

The research project *Democracy – Technological Transformation – Society (DeTS)* examines the influence of AI-co-moderated language on liberal democracy, specifically investigating how AI may reinforce or undermine democratic processes. Funded and coordinated by the Jagiellonian University, the project is conducted by a consortium comprising four European universities and two non-academic research organizations. The project is scheduled to run for two years, commencing in November 2024, with a primary focus on Europe and the European Union.

The third final DeTS report synthesizes and extends earlier project findings[1] by compressing the AI–democracy problem space into **three analytically sufficient pillars**: (I) **Public Arena & Gatekeeping**, (II) **Trust & Epistemic Agency**, and (III) **Capabilities & Governance**. Building on DeTS Report I's diagnosis of risks and opportunities of AI-co-moderated language for liberal democracy and its emphasis on principles and multi-actor governance, the report integrates DeTS Report II's systemic account of algorithmic gatekeeping, information-chain transformation, and the erosion of epistemic common ground. We propose a compact mechanism narrative linking AI-mediated visibility and persuasion to democratic outcomes (pluralism, legitimacy, participation). Moreover, we outline a minimal portfolio of interventions that combines stakeholder-specific AI literacy with institutional safeguards

[1] see the first report, issued on 22/01/202, Declich and Feudo 2025, available here and the second one issued on 30/07/2025, Declich et al. 2025, available here).

aimed at preserving pluralism while improving assessability and accountability. The report concludes with a focused future research agenda aligned to each pillar, designed to support comparative, mixed-methods inquiry and policy-relevant evaluation across the European context.

From DeTS Diagnosis to a Three-Pillar Model

1. Objectives

1.1 Purpose and contribution

DeTS investigates how **AI-co-moderated language** can strengthen or weaken liberal democracy and why the issue demands ongoing research under rapid technological and social change. DeTS Report I framed the problem around AI-enabled distortions of the information environment (mis/dis/malinformation, deepfakes, computational propaganda) and highlighted the difficulty of designing governance that mitigates harm without inadvertently reducing democratic pluralism. DeTS Report II extended this by shifting attention from "downstream" users to the **entire information chain** and the **public arena**, emphasizing algorithmic gatekeeping, assessability, and the erosion of epistemic agency under generative AI.

Report III offers an academic-facing synthesis that **reduces complexity without oversimplifying**: it proposes three pillars that capture the main causal channels through which AI affects democratic life and that can guide both intervention design and future research.

1.2 The three pillars and the analytical thesis

A brief overview of the three pillars is presented below.

Pillar I— Public Arena & Gatekeeping:

AI intensifies the transition from journalistic gatekeeping to platform- and AI-mediated selection, reshaping visibility, agenda-setting, and the conditions for democratic contestation.

Pillar II— Trust & Epistemic Agency:

Generative AI increases epistemic uncertainty by scaling persuasive content not designed for truth, undermining citizens' capacity to form political will on shared grounds.

Pillar III— Capabilities & Governance:

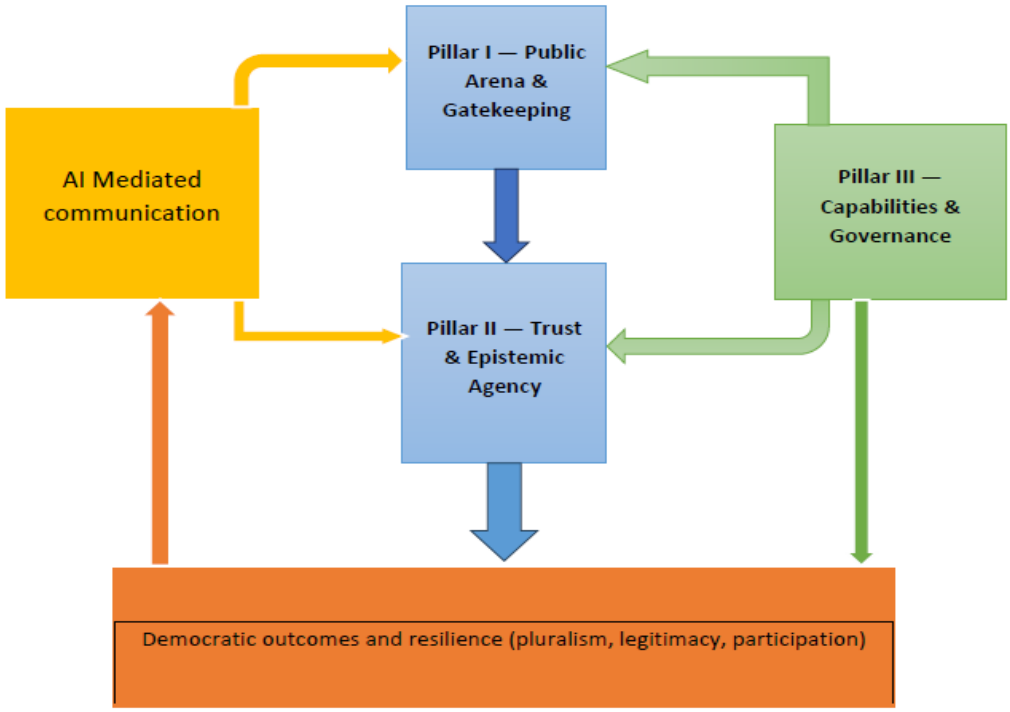
Democratic resilience requires the following two conditions: stakeholder-specific AI literacy and institutional safeguards. In this framework, governance must avoid "backfire" effects such as over-blocking and reduced pluralism.

We argue that **democratic resilience in the AI era** depends on how:

- **capabilities and governance** (Pillar III) are built so to influence how:
    - the public arena is kept **assessable** (Pillar I) (scrutinizable) despite AI mediation, and
    - **trust and epistemic agency** (Pillar II) are preserved

In this way, manipulation operated on Pillar I and II by AI mediated communication could be tamed while maintaining pluralism and legitimate expression.

Figure 1: The Analytical Thesis:



2. Pillar I: Public Arena & Gatekeeping (System Layer)

2.1. Why the public arena is the critical unit of analysis

DeTS Report II conceptualizes the **public arena** as the set of societal functions that make society visible to itself, enable contestation, and sustain shared and counterpublic spaces where identities and preferences are formed. In this view, democratic risks emerge not only

from false content but from transformations in how content becomes visible and authoritative—i.e., from changes in the socio-technical processes of gatekeeping.

2.2. The gatekeeping shift under AI

AI accelerates an ongoing shift from institutionalized journalistic gatekeeping to **algorithmic gatekeeping/selection** mediated by platforms and AI tools. This shift matters democratically because it can (a) redistribute agenda-setting power to intermediaries that are hard to regulate democratically, and (b) alter pluralism by privileging scale, engagement metrics, or "average" patterns over minority visibility.

2.3. Assessability as a democratic condition

A central risk is the weakening of **assessability**—the ability of citizens, researchers, and institutions to scrutinize how the public arena operates. When AI systems and platforms are opaque (in data, training, or selection logics), debate cannot evaluate whether public visibility is being shaped in ways compatible with democratic norms. This pushes DeTS from content-focused concerns toward infrastructure-focused concerns: democracy requires not only speech, but the assessable conditions under which speech competes.

2.4. Minimal implication for intervention design

Pillar I implies that resilience strategies must include measures that **improve institutional and infrastructural transparency**, not merely individual discernment. In practice, this means: strengthening verification and provenance workflows in news ecosystems and developing scrutiny expectations for AI intermediaries that shape public visibility.

3. Pillar II: Trust & Epistemic Agency (Citizen Layer)

3.1. Epistemic agency and democratic self-rule

DeTS Report II foregrounds **epistemic agency**—the degree of control individuals have over belief formation and revision—as a prerequisite for democratic will-formation. The AI challenge is not limited to discrete episodes of misinformation; it includes the systemic possibility that citizens lose confidence in their ability to know what is real, thereby weakening participation, deliberation, and legitimacy.

3.2. Why generative AI raises epistemic uncertainty

Generative AI systems produce fluent, convincing outputs but are not intrinsically oriented toward truth; at scale, this can erode an "epistemic common ground" necessary for liberal democratic discourse. The consequence is a climate of uncertainty in which correction becomes harder, the burden of verification shifts to individuals, and institutional trust can be replaced by generalized scepticism ("nothing is reliable").

3.3. Fragmentation dynamics

Algorithmic and AI-mediated personalization can deepen fragmentation by producing distinct informational environments across groups, complicating democratic compromise and shared problem-solving. This dynamic interacts with Pillar I: gatekeeping structures shape what people can collectively deliberate about, while erosion of epistemic agency shapes whether people believe deliberation is meaningful.

3.4. Minimal implication for intervention design

Pillar II implies that resilience cannot be reduced to detection technologies or "spot the fake" training. Instead, it requires (a) civic verification habits, (b) institutional reference points that lower the cognitive burden on individuals, and (c) treating epistemic agency as a measurable outcome for evaluation over time.

4. Pillar III: Capabilities & Governance (Action Layer)

4.1. Why capabilities must be stakeholder-specific

DeTS Report II argues that AI literacy should not be framed solely as a workforce "skills gap." It is a civic and institutional capability that must vary by role: citizens, media professionals, professional communities (including scientific communities), public entities, civil society organizations, and system owners face different responsibilities and risks. This aligns with DeTS Report I's insistence that solutions require cooperation among heterogeneous actors with different value frameworks (rights, innovation, economic interests, democratic principles).

4.2. A minimal capability architecture (four groups)

To keep the model compact, we propose the following capability targets:

a. **Citizens/voters:** provenance awareness, verification routines, and critical interaction with AI outputs.

b. **Media professionals:** responsible newsroom use of AI, verification/provenance workflows, disclosure norms.

c. **Professional communities (including scientific communities):**awareness of the societal and professional implications of AI related risks, procurement literacy, verification routines, and critical interaction with AI outputs.

d. **Public sector & civil society organizations:** oversight competence, procurement literacy, accountability tools and redress mechanisms.

4.3. Governance trade-offs: protecting democracy without shrinking it

DeTS Report I warns that mitigation approaches (guardrails, detection, moderation, labelling) can generate unintended democratic harms, including restricting legitimate expression and limiting pluralism – especially in sensitive political periods. This risk is reinforced by Report II's public arena perspective: interventions that alter visibility can reshape democratic contestation itself. Therefore, governance must include safeguards – transparency of criteria, proportionality, human oversight where appropriate, and contestability/appeal pathways – to avoid "over-blocking" and legitimacy loss.

4.4. Minimal implication for intervention design

Pillar III supports a combined approach: deploy stakeholder-first literacy as lifelong learning, institutionalize multi-actor cooperation (research – media – policy – civil society – tech), and evaluate governance tools explicitly for pluralism impacts, not only for harm reduction.

5. Conclusions, Recommendations, and Focused Future Research

5.1. Synthesis

The three pillars jointly explain how AI affects democratic life: AI reshapes the public arena through gatekeeping and assessability (PI), alters the epistemic conditions of citizenship (PII), and demands capability-building and governance that prevent manipulation without shrinking pluralism (PIII). This compression preserves the main conceptual advances of DeTS while providing a tractable framework for policy-relevant research.

5.2. Six recommendations (academic-facing, implementation-neutral)

To make this approach practical, the following recommendations are formulated for possible policy initiatives.

a. **Operationalize assessability** as a normative requirement for AI-mediated public visibility (define what scrutiny should be possible).

b. **Strengthen accountable gatekeeping capacity** (verification and provenance standards) in media ecosystems exposed to AI competition.

c. **Monitor epistemic agency and trust** as core democratic outcomes under AI-mediated communication.

d. **Deploy stakeholder-specific AI literacy** (citizens, media, professional communities, public/civil oversight) as a repeatable, lifelong-learning practice.

e. **Evaluate governance tools for pluralism impacts** (avoid over-blocking; ensure contestability and transparency).

f. **Institutionalize multi-actor cooperation** so that governance and literacy evolve with technology and social practice.

5.3. Three future research paths (one per pillar)

R1 (Pillar I): Public Arena Audit Programme. Comparative audits of AI-mediated gatekeeping and assessability standards across national contexts and platforms/news ecosystems.

Future research should establish a comparative audit programme that makes AI-mediated gatekeeping empirically visible and comparable across countries, platforms, and news ecosystems. The purpose is to document how selection and ranking systems distribute and redistribute agenda-setting practices and influence pluralism, and to translate the report's key normative requirement (like assessability) into concrete, testable benchmarks. This path may produce repeatable public arena audits (e.g., periodic tests around elections or states of different crises) that track where and how visibility is managed, whether minority and counterpublic voices are heard or silenced, and what forms of scrutiny (data access, provenance signals, explanation of criteria, contestability) are possible for researchers and institutions. In practical terms, this should end in something policymakers can actually use: a handful of indicators and a yearly record that makes it easy to see where the public arena is losing assessability – and where attention is most urgently needed.

R2 (Pillar II): Epistemic Agency Observatory. Longitudinal monitoring of trust, verification capacity, and epistemic common ground under increasing AI content saturation.

A second research path should build a longitudinal observatory focused on trust and epistemic agency under conditions of rising AI-generated content. The report's diagnosis is that generative AI scales persuasive language not only and exclusively designed for truth, which increases epistemic uncertainty, raises individual costs of verification, and erodes the shared reference points needed for democratic formation of political will. This calls for systematic monitoring of (a) citizens' perceived ability to know what is reliable, (b) their practical verification practices and routines and (c) the degree of epistemic common ground across groups in a fragmented information environment. The observatory should combine repeated population measurements with targeted qualitative tracking of real verification practices, producing early-warning signals (e.g., sudden drops in confidence, rising generalized scepticism) and a stable measurement toolkit that treats epistemic agency as a democratic outcome. Such evidence would directly inform which institutional reference points and verification supports reduce cognitive burden without pushing citizens into cynicism or disengagement leading to antipolitical actions/behaviors/habitats.

R3 (Pillar III): Capability & Governance Trials. Mixed-method evaluation of literacy interventions and governance safeguards, explicitly testing trade-offs between harm reduction and pluralism (check monopolistic take-over of AI systems). Evaluation should be conducted in relation to the development of the **crisis of democracy as a social process,** using the results of the ongoing research on the matter.

A third path may evaluate, in realistic settings, which combinations of stakeholder-specific AI literacy and governance safeguards, improve resilience without shrinking democracy. The report emphasizes that capabilities must differ by role (citizens, media professionals, professional/scientific communities, and public sector/civil society oversight), and that governance can fail through over-blocking, loss of contestability, or reduced pluralism - especially during sensitive/tense political periods. Research should therefore run mixed-method tests of literacy modules, newsroom source credibility and disclosure norms. At the same time, governance tools should be tested (labelling, moderation criteria, transparency duties, appeal pathways) against a two-sided standard. Importantly, this agenda should also examine structural risks of dependency and concentration, including the possibility of a monopolistic take-over of AI systems, by studying lock-in dynamics in media and public-sector funding and identifying safeguards that preserve institutional autonomy. Evaluation should be placed in the broader, ongoing research on the crisis of democracy as a social process, so that observed effects are interpreted as part of a longer-term and complex development of the 21st century societies.

5.4. Closing statement

The findings of the DeTS project indicate that the relationship between artificial intelligence and liberal democracy cannot be adequately framed as a set of isolated technological risks requiring technical fixes. Rather, what is at stake is a gradual but consequential reconfiguration of the communicative environment within which democratic life unfolds. AI-mediated language does not simply add new content to the public sphere; it increasingly shapes the conditions under which information becomes visible, credible, and politically actionable. In this sense, the issue is less about individual instances of misinformation than about the evolving socio-technical infrastructure through which public knowledge and collective judgement are formed.

The three-pillar framework developed in this report highlights that democratic resilience in AI-saturated environments depends on preserving the assessability of public-arena dynamics, safeguarding citizens' epistemic agency, and strengthening role-specific capabilities across institutional and societal actors. From a policy standpoint, this implies the need to move beyond content-focused interventions toward governance approaches attentive to how visibility, authority, and trust are structured by algorithmic systems. At the same time, regulatory responses (such as moderation regimes, automated detection tools, or labelling mechanisms) must be designed with caution, as measures intended to enhance informational integrity may inadvertently constrain legitimate expression or weaken pluralism, particularly in politically sensitive contexts.

Ultimately, the AI–democracy nexus should be understood not as a bounded regulatory problem, but as an ongoing transformation of democratic communication itself. Addressing it will therefore require cumulative research, adaptive governance, and sustained cooperation across media, policy, research, and civil society. A full bibliography supporting this report is provided in DeTS Reports I and II.

References

Declich A., Feudo F. (2025), Artificial Intelligence and Democracy (Democracy – Technological Change – Society Project (DeTS), Report 1, Jagiellonian University in Krakow, K&I – Conoscenza e Innovazione
https://ism.uj.edu.pl/documents/3337228/147938651/Raport+DeTS/01416ba2-1376-45ac-ab8f-8643426b11b6

Declich A., Feudo F., Banaś M., (2025)AI and Democracy: Navigating Digital Literacy Amid the Rise of AI-generated News, Report II: Democracy – Technological Change – Society (DeTS), Jagiellonian University in Krakow, K&I – Conoscenza e Innovazione, https://fudelab.id.uj.edu.pl/publications/-/journal_content/56_INSTANCE_6OhNUsv6IRLF/152110637/159279449

Further bibliographic references are provided in DeTS Report I and DeTS Report II.

by Fabio Feudo & Andrea Declich
Knowledge & Innovation
In cooperation with
Monika Banaś
Jagiellonian University