

KT4D

Knowledge Technologies
for Democracy

POLICY BRIEF

The Evolving Role of Trust in AI-Mediated Societies

Authors

Trinity College Dublin

Jennifer Edmond

Fundación Cibervoluntarios

María Ruíz de Assin de los Santos

Demos Helsinki

Atte Ojanen

Strane Innovation

Tiffany Morisseau

Warwick University

Jessica Sutherland

Trust-IT Services

Gabriela Rogowska

Gabriele Quattrocchi

Graphic Design

Gianluca Savini e Andrea Greco - Trust-IT Services

Organisations involved in KT4D



Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin



BEYOND
THE HORIZON

CIBER
VOLUNTARIOS.org



COMMpla
Communication Platforms
and Online Solutions

DEMOS
HELSINKI



HYBRID CORE



IRMiR
INSTITUTE OF
URBAN AND REGIONAL
DEVELOPMENT

Strane
Innovation



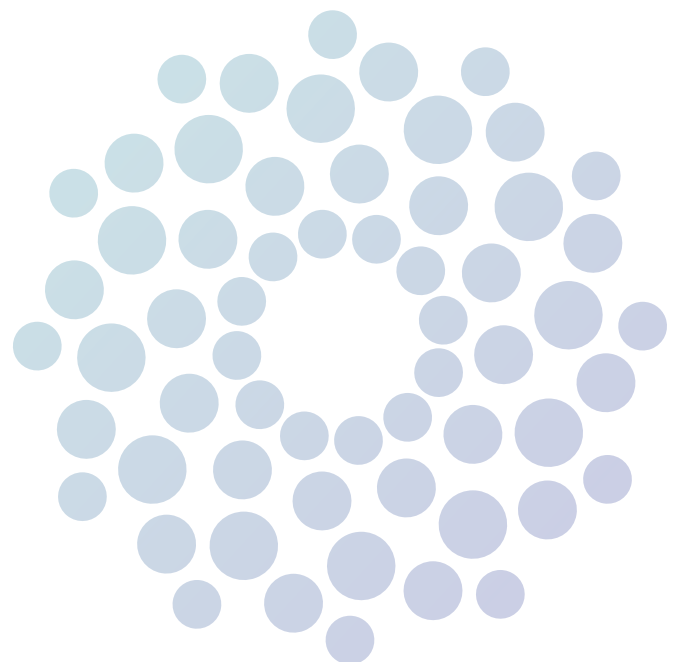
Trust-iT services
communicating to markets



WARWICK
THE UNIVERSITY OF WARWICK

Table of contents

Executive Summary.....	2
1. AI as a new actor in the trust equation.....	3
1.1 Societal Challenges.....	3
1.2 Policy Challenges	5
2. Recommendations	6
Recommendation 1: Ensure the traceability and authenticity of information by public institutions	6
Recommendation 2: Encourage organisations to communicate on shared rules of AI use to sustain social trust.....	6
Recommendation 3: Foster contextualised critical digital literacy programmes	7
Recommendation 4: Ensure AI governance is responsive to interculturality and cultural differences	8
Recommendation 5: Establish mandatory human oversight for all AI systems involved in critical decision-making, particularly in employment, education, healthcare, and legal contexts.	9
Recommendation 6: Ensure epistemic responsibility in online knowledge dissemination through increased scrutiny and regulation of online platforms	10
Recommendation 7: Adopt broader, infrastructural lens to democratic AI governance.....	10
Recommendation 8: Ensure sufficient capacities to enforce current AI regulation.....	10
3. About KT4D Project	11



Executive Summary

Trust in democratic institutions is being fundamentally redefined as Artificial Intelligence increasingly mediates information exchange and influences decision-making processes. The KT4D project examines how artificial intelligence affects trust at multiple levels: between citizens and institutions, across different cultural contexts, and within organisations where Artificial Intelligence is deployed without clear standards.

The policy brief presents eight recommendations grouped across institutional, organisational, educational, and governance dimensions. At the institutional level, public bodies should ensure traceability and authenticity of information through authenticated digital signatures and origin labels. Organisations deploying AI systems must establish and communicate shared rules for AI use to keep interpersonal and institutional trust.

Educational recommendations emphasise contextualised critical digital literacy programmes that reflect local languages, social norms, and cultural values. Governance recommendations call for AI frameworks responsive to intercultural differences, mandatory human oversight for critical decision-making systems, epistemic responsibility in online knowledge dissemination, infrastructural approaches to democratic AI governance, and sufficient enforcement capacities for current regulation.

The central policy challenge is that while the European Union has addressed the effects of artificial intelligence through regulations, the focus has recently shifted toward innovation and deregulation. This creates tension between upholding democratic processes and promoting the rapid deployment of artificial intelligence, as this advanced knowledge technology can indirectly disrupt democracy in ways that static risk categories fail to capture.



1. AI as a new actor in the trust equation

Trust is essential in democratic societies, forming a precondition for citizens to rely on institutions and engage in decision-making. Yet, trust is being redefined as Artificial Intelligence (AI) increasingly mediates information exchange, and affects decision-making processes and public discourses. The influence of algorithms and the rapid dissemination of news and information across cultural contexts

make it increasingly difficult to identify trustworthy sources of knowledge. As of today, trust depends not only on institutional integrity but also on digital literacy and technical transparency. Understanding how trust is built, maintained, and sometimes eroded in technologically mediated environments is central to safeguarding critical thinking and democratic participation.

1.1 Societal Challenges

What does it mean to receive, understand, or produce information in AI-mediated environments? The growing likelihood that any given piece of information, image or video has been generated or transformed by AI obfuscates how citizens assess and relate to information. It is becoming increasingly complex to identify in what context information was produced and the intentions underpinning it. The uncertainty caused by this opacity can (sometimes intentionally) undermine citizens' trust in institutions, which are less able to serve as reliable reference points. Therefore, maintaining trust depends on strong collective mechanisms ensuring that information shared by institutions can be reliably traced, verified, and contextualised, especially on issues that truly matter to citizens.

The intertwining of trust in institutions and trust in AI

The [2025 Edelman Trust Barometer](https://www.edelman.com/trust)¹ shows a clear trend: people who feel a stronger sense of grievance — those who believe that the system has failed or excluded them — tend to trust artificial intelligence less and feel less comfortable with its use by businesses. This decline in confidence reflects more than technological skepticism; it points to a deeper decay of institutional and social trust. This helps explain the 22-point drop in AI trust among those with high grievance in the survey. For many, the technology is not viewed as a tool for transparency, but rather as another source of informational uncertainty.

1 Edelman. (2025). 2025 Edelman Trust Barometer. Edelman. <https://www.edelman.com/trust>



With Greater Grievance, More Suspicious of Artificial Intelligence

Percent who say, by sense of grievance

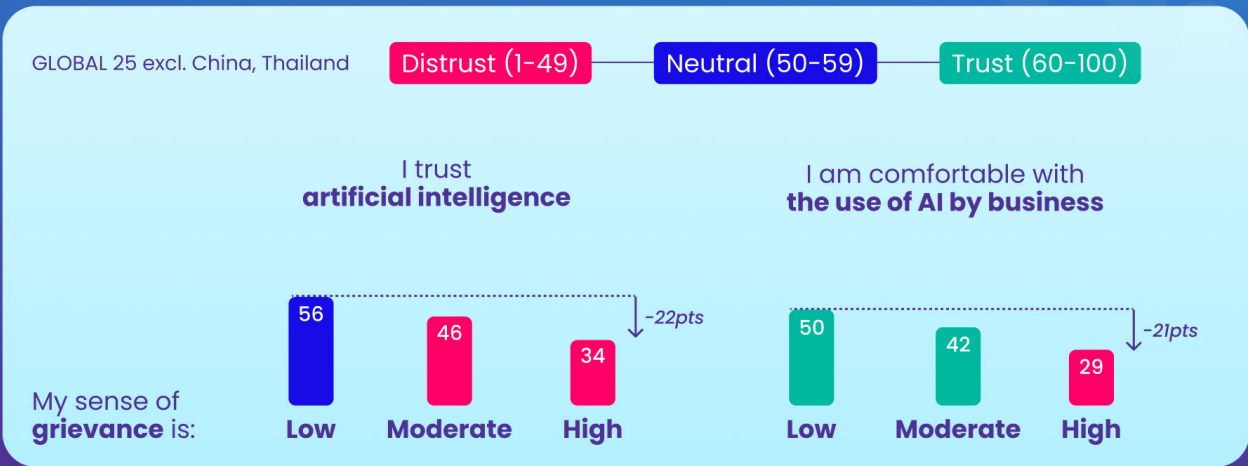


Figure 1. The figure adopted from the Edelman Trust Barometer (2025).

The role of culture in trust mediation

Attitudes toward AI and institutional trust are also culturally mediated. Depending on whether a country has high institutional trust or not, citizens may view AI as a tool for efficiency and public benefit, or exclusion and manipulation. There is also a clear trust gap between the Global North and South. Citizens in Western democracies, such as the US, UK, Germany, and Japan, generally show low and declining trust in AI (37–47%). In contrast, many countries in the Global South, particularly in Asia and the Middle East (China, India, UAE, Indonesia, Saudi Arabia), exhibit high and stable trust in the technology (70–77%). This pattern suggests that trust is not merely about elections or democratic participation but shaped by broader structural, social, cultural, and governance factors, creating a North–South divide.

Trust within organisations

Another challenge concerns how AI affects trust within organisations. The development of AI in

collective and institutional settings has indeed direct effects on interpersonal trust between members of communities, whether corporate, scientific, or educational.. Beyond individual interactions, AI affects how these communities negotiate shared norms. When the use of AI is introduced without clear standards, it can generate uncertainty not only linked to the technology itself, but also about people’s intentions, competencies and responsibilities. The absence of a shared understanding of what is an acceptable or legitimate use of AI may erode the mutual expectations underlying cooperation. While there are some generally shared intuitions on this issue (for example, the use of AI is considered more legitimate for administrative support than hiring decisions), establishing explicit and acceptable frameworks for the use of AI is therefore a social and democratic necessity. (For related insights, see KT4D Social Risk Toolkit Module B: AI, trust and awareness, “AI and the valuation of human work” DOI [10.5281/zenodo.17977403](https://doi.org/10.5281/zenodo.17977403)).



1.2 Policy Challenges

The European Union (EU) is tackling AI's societal effects through a multifaceted set of policies, with implications for institutional trust. On the regulatory front, laws like the GDPR, the Digital Services Act (DSA), and the Digital Markets Act (DMA) govern data processing and the behaviour of major online platforms where disinformation often spreads. The EU AI Act (Regulation (EU) 2024/1689) addresses democratic threats more directly by classifying systems that influence voting as “high-risk,” while new democracy-focused policies like the European Democracy Shield [Brussels, 12.11.2025 JOIN(2025) 791 final - European Democracy Shield: Empowering Strong and Resilient Democracies] are designed to combat foreign information manipulation and protect media pluralism.

The central policy challenge, however, is that AI can also indirectly disrupt democracy in ways that static risk categories fail to capture. At the same time, the EU focus has shifted towards innovation and deregulation, as evident in the [Digital Omnibus](#)² and the AI Continent Action Plan³. The latter includes initiatives like the Apply AI Strategy, the Cloud and AI Development Act, and the development of AI Factories. While digital sovereignty and public AI infrastructure is certainly needed to build trust, one can also identify a tension between upholding democratic processes and promoting rapid, profit-driven AI deployment across society.

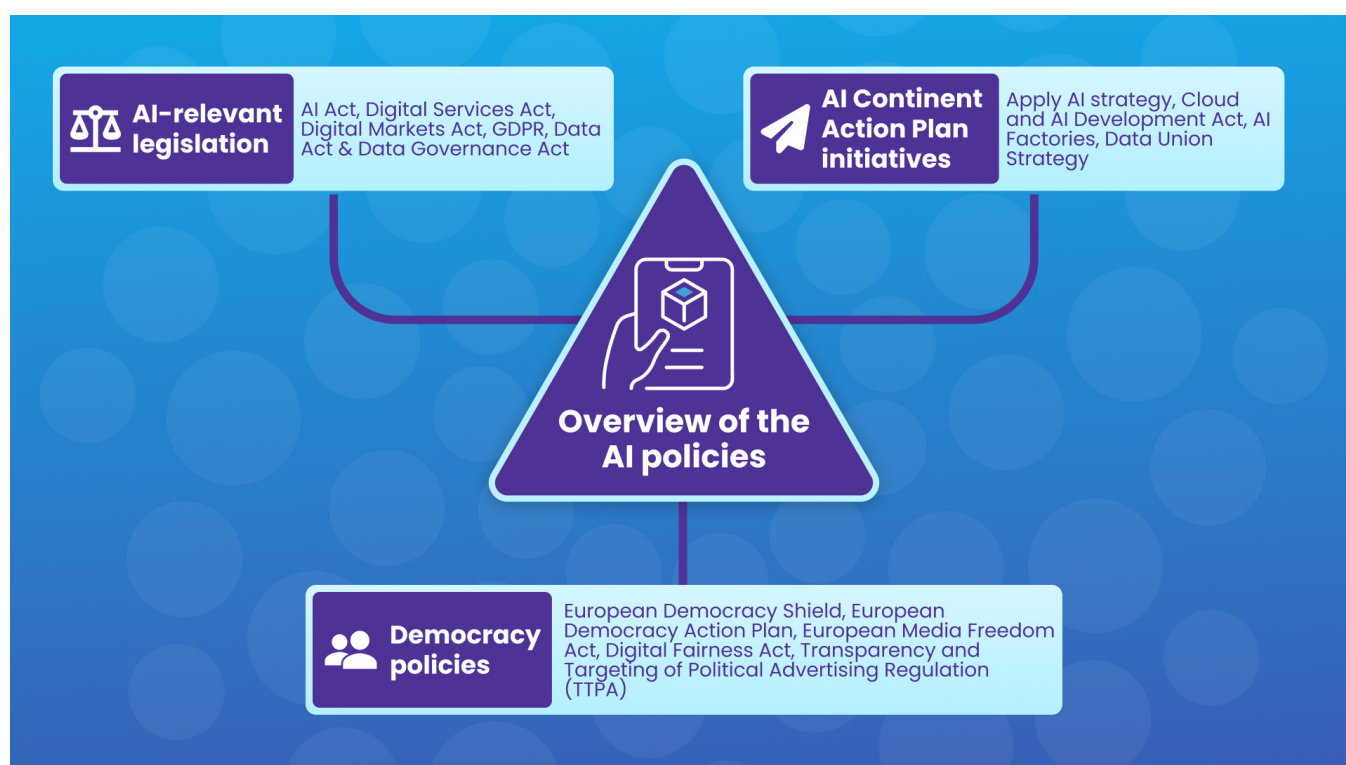


Figure 2. Overview of the AI-related policies.

² Brussels, 19.11.2025 COM(2025) 836 final 2025/0359 (COD) Proposal for a Regulation Of The European Parliament and of the Council amending Regulations (EU) 2024/1689 and (EU) 2018/1139 as regards the simplification of the implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI).

³ Brussels, 9.4.2025 COM(2025) 165 final Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions, AI Continent Action Plan.



2. Recommendations

☑ Recommendation 1: Ensure the traceability and authenticity of information by public institutions

Public institutions should establish systems that guarantee the traceability and authenticity of the information they produce and disseminate (for example, through authenticated digital signatures or institutional origin labels certifying that the content comes from an identifiable public body). These mechanisms should include a visible and easily understandable sign or label, allowing citizens to recognise genuine institutional communication.

🔍 Why it is important

As AI-generated content is growing exponentially, citizens face growing uncertainty about the origin and reliability of information. Ensuring that information coming from public bodies can be clearly identified and traced back to their legitimate source is therefore essential to maintain trust and to prevent the manipulation of public discourse.

☑ Recommendation 2: Encourage organisations to communicate on shared rules of AI use to sustain social trust

Each organisation that deploys AI systems (whether in public administration, research, education, business, or civil society) should establish and make available its own shared rules for AI use, such as universities' guidelines for the use of generative AI. Ideally these are discussed internally and informed by a clear mapping of how AI is used across the organisation's activities.

🔍 Why it is important

When organisations adopt AI tools without clear internal standards, uncertainty arises not only about the technology but also about human intentions, competencies, and responsibilities. Collaborators and citizens need to know when and how AI is expected to be used. Without this shared understanding, interpersonal and institutional trust may be undermined.



✓ Recommendation 3: Foster contextualised critical digital literacy programmes

Design and implement contextual critical digital literacy programmes within local institutions (schools, libraries, and civil society organisations) that integrate and embed social, linguistic, and political specificities of each community. These should include critical consumption skills, including analysis and awareness.

🔍 Why it is important

Digital and AI literacy cannot be universalised; it must reflect local languages, social norms, and cultural values. A culturally contextualised approach ensures that citizens understand and trust the digital tools that shape democratic participation.

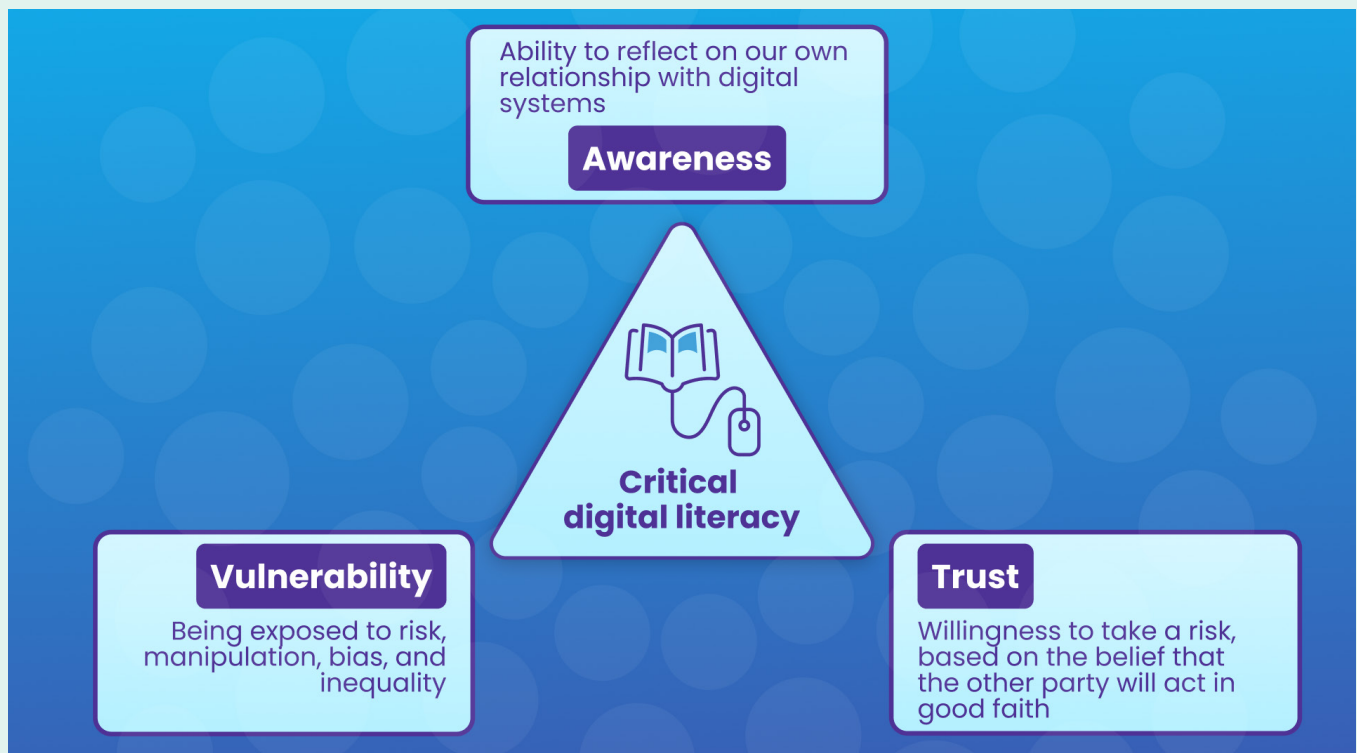


Figure 3. Illustration of the components of critical digital literacy: awareness, trust and vulnerability

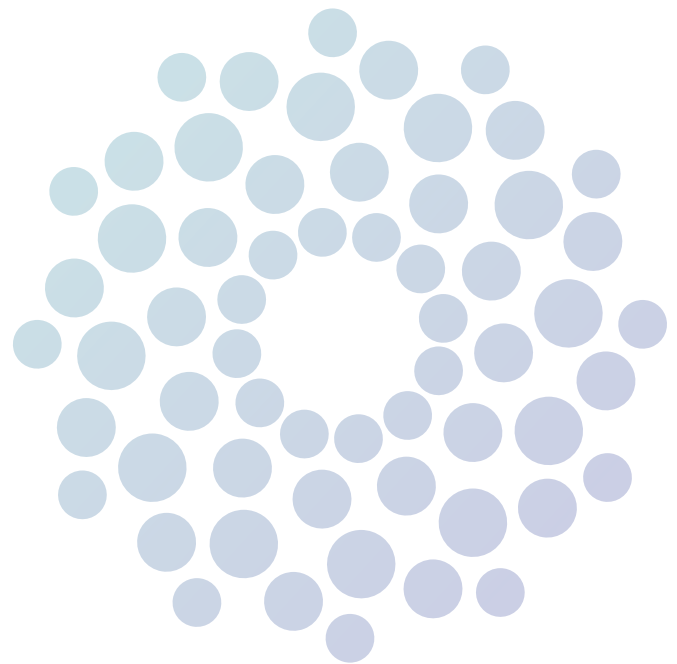


☑ Recommendation 4: Ensure AI governance is responsive to interculturality and cultural differences

Utilise United Nations forums like Global Digital Compact⁴, the Global Dialogue on AI Governance⁵ and the Independent International Scientific Panel on AI⁶ to incorporate intercultural perspectives to AI governance. Ensure that all regions, communities and cultures are heard about their views on trust, risk, and the role of technology to shape AI policies from the start to better align it with the needs of the people it aims to serve.

🗨 Why it is important

AI now affects countries and communities across the world, but many are still not represented in international discussions. Intercultural dialogue is crucial for understanding how different communities conceptualise trust, risk and authority. Without this dialogue, AI governance risks being shaped by narrow cultural assumptions that may not translate across contexts.



4 The Global Digital Compact is a comprehensive global framework for digital cooperation and governance of artificial intelligence. Twenty years after the World Summit on the Information Society, it charts a roadmap for global digital cooperation to harness the immense potential of digital technology and close digital divides. On 22 September 2024, world leaders convened in New York for the Summit of the Future, where they adopted a Pact for the Future that includes a Global Digital Compact.

5 The Global Dialogue on AI Governance provides an inclusive platform within the United Nations for states and stakeholders to discuss the critical issues concerning AI facing humanity today.

6 Established with Resolution A/RES/79/325 on 26 August 2025, the Independent International Scientific Panel on AI serves as the first global scientific body on Artificial Intelligence (AI), bringing together leading experts to assess how AI is transforming society and economy.



☑ **Recommendation 5: Establish mandatory human oversight for all AI systems involved in critical decision-making, particularly in employment, education, healthcare, and legal contexts.**

Implement a Democracy-in-the-Loop framework that genuinely enables the public and affected communities to shape how ML systems are developed and governed throughout their lifecycle. Transparency in this context means ensuring information is understandable and actionable for those it affects, rather than simply making technical documentation available. Individuals must have robust mechanisms to challenge AI-assisted decisions and seek redress when systems fail them. This means addressing the power asymmetries that ML systems create or amplify by requiring impact assessments and by providing resources for marginalised communities to participate in governance processes.

🗨️ **Why it is important**

AI and automated decision-making systems are becoming increasingly prevalent in essential services. Whilst this may lead to cost and efficiency savings, to foster public trust and confidence in AI-assisted decisions - and ultimately in our democratic institutions - requires that human oversight and accountability must remain central. Existing frameworks, such as human-in-the-loop and human-on-the-loop provide some level of accountability, however more must be done to protect and bolster our democratic and cultural values, as key pillars of trust in institutions.

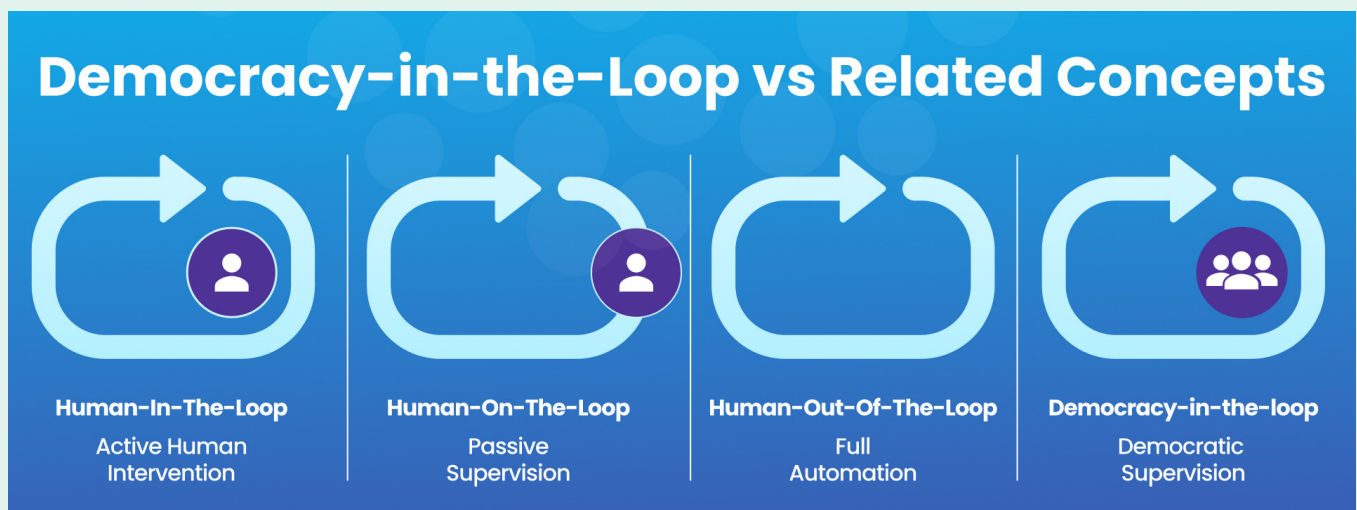


Figure 4. Illustration of the Democracy-in-the-Loop and Related Concepts



☑ Recommendation 6: Ensure epistemic responsibility in online knowledge dissemination through increased scrutiny and regulation of online platforms

Scrutiny of AI-driven algorithms should ensure that these platforms promote accurate and verifiable information, rather than content that is misleading or inaccurate. The EU ought to enforce the Digital Services Act and Digital Markets Act in a robust way to steer the behaviour of large platforms that act as epistemic gatekeepers.

🔔 Why it is important

Knowledge dissemination has increasingly moved online to search engines and social media platforms. Those who control these platforms are therefore mediators of epistemic trust and have immense power in shaping what knowledge is disseminated and how.

☑ Recommendation 7: Adopt broader, infrastructural lens to democratic AI governance

Leverage different policies like the AI Act, Digital Services Act, Digital Markets Act and European Democracy Shield in harmony to tackle concentration of power in AI at the infrastructural layer. Incentivise investments into digital public infrastructure for a sovereign, democratically governed European AI stack.

🔔 Why it is important

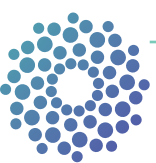
Focussing merely on AI systems deployment disregards the importance of decision-making in the design and development processes. Democratic governance and trust are difficult to realise afterwards if the digital public sphere is owned by large private platforms and gatekeepers with profit incentives.

☑ Recommendation 8: Ensure sufficient capacities to enforce current AI regulation

Invest into public sector capacities for implementation and enforcement of current AI and data regulation to provide legal certainty. Make robust enforcement of democratic safeguards a tangible competitive advantage that distinguishes the EU from unreliable, authoritarian countries by providing sufficient resources to national authorities and the EU AI Office.

🔔 Why it is important

The EU has established a strong digital rulebook to protect fundamental rights, but these rules are only effective if they are properly enforced. Haphazardly implemented [Digital Omnibus](#) risks undermining citizens' trust in these safeguards.



3. About KT4D Project

The Knowledge Technologies for Democracy (KT4D) project offers a distinctive approach to the growing tensions between AI and big data technologies, and democratic processes. We do this by placing cultural values and identity formation at the heart of our understanding.

The KT4D project uniquely positions technologies such as AI and big data as advanced knowledge technologies (AKTs), addressing the challenge of integrating a cultural perspective into the study of technology and democracy.

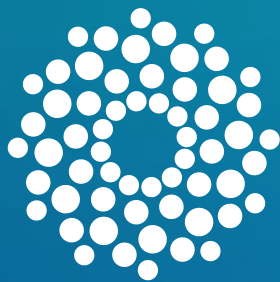
The project recognises that AKTs have become fundamentally interwoven into our societies, our cultures and indeed into our expectations and conceptions of democratic governance and exchange.

The project recently released the KT4D Toolkit that collects all KT4D results in one place, empowering stakeholders with knowledge, methods, and tools for democratic resilience in the AI era.

Want to comment on this policy brief? You can do so here

kt4democracy.eu





KT4D

Knowledge Technologies
for Democracy

Follow the KT4D project

 kt4democracy.eu

 [@kt4democracyproject](https://www.youtube.com/@kt4democracyproject)

 [/company/kt4democracy/](https://www.linkedin.com/company/kt4democracy/)

 [@KT4Democracy](https://twitter.com/KT4Democracy)

 [@kt4democracy.bsky.social](https://bsky.social/kt4democracy)

 [/communities/kt4democracy](https://zenodo.org/communities/kt4democracy)



Funded by
the European Union