Project Number: 101094302 Start Date of Project: 01/02/2023 Duration: 36 months

Deliverable 5.1

Framework for Democratic AI Governance

Dissemination Level	PU
Due Date of Deliverable	31/01/2026, M36
Actual Submission Date	31/01/2026
Work Package	WP5 | Trustworthy Software Development and Regulation
Task	T5.1 | Craft a governance framework and a roadmap for democratic use of AI
Type	Report
Version	V2 Final
Number of Pages	p.1 – p.56

Deliverable Abstract

This deliverable presents the KT4D's framework for democratic AI governance. The framework seeks to conceptually clarify the different interpretations of AI governance in and for democracy, in light of six pillars of democracy and the different governance mechanisms along the AI lifecycle. This updated version of the deliverable has been complemented by a more concrete policy roadmap and recommendations to guide EU policymakers in governing AI democratically.

Funded by the European Union

KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

DELIVERY SLIP

Date	Name	Partner/Activity	Date
Moderated by:	Atte Ojanen	DEMOS	10/12/2024
Reviewed by:	Eva Power Tiffany Morisseau Eleonora Lima	TCD STRANE TCD	16/12/2024 22/01/2026 22/01/2026
Approved by:	Eva Power	TCD	31/01/2026

DOCUMENT LOG

Issue	Date	Comment	Author	ORCID ID
v0.5	10/12/2024	First Draft	Atte Ojanen Anna Björk Johannes Anttila Anna Mäkituomas Kate Francis	0000-0002-7543-5821 0000-0002-2008-0920 0000 0003 2830 5183
v0.7	16/12/2024	Peer Review	Eva Power Tiffany Morisseau	0009-0000-5118-0417 0000-0002-2523-7755
v.1.0	19/12/2024	Submitted version 1, with reviewer comments addressed	Atte Ojanen	0000-0002-7543-5821
v1.5	10/01/2026	Second Draft	Atte Ojanen Anna Björk Vera Djakonoff Sahib Singh	0000-0002-7543-5821 0000-0002-2008-0920 0000-0002-8715-6686 0000-0003-1213-7453
v.1.7	22/01/2026	Peer Review	Eleonora Lima Tiffany Morisseau	0000-0001-7578-8005 0000-0002-2523-7755
v2.0	30/01/2026	Final version	Atte Ojanen	0000-0002-7543-5821

TERMINOLOGY

Terminology/Acronym	Definition
EU	European Union
AI	Artificial intelligence
ADM	Automated Decision Making
AI Act	Artificial Intelligence Act; Regulation (EU) 2024/1689
GDPR	General Data Protection Regulation; Regulation (EU) 2016/679
DSA	Digital Services Act; Regulation (EU) 2022/2065
DMA	Digital Markets Act; Regulation (EU) 2022/1925
GPU, TPU	Graphics Processing Unit & Tensor Processing Unit
API	Application Programming Interface
KYC	Know Your Customer
CoE	Council of Europe
UNESCO	United Nations Educational, Scientific and Cultural Organization
GPAI	Global Partnership on Artificial Intelligence
OECD	Organisation for Economic Co-operation and Development

Funded by the European Union

KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

Table of Contents

Introduction 2
1. Values behind European AI and technology regulation 3
    1.1. Development of the EU digital policy and legislation: from embedded ethics to a risk-based approach 3
    1.2. Shifting AI policy priorities 5
    1.3. Beyond the EU: examples of legislative initiatives and guidelines 7
2. Pillars of democracy 9
    2.1. Impact of AI on democratic pillars 9
    2.2. Intersection of the pillars in relation to AI 13
3. AI governance across the lifecycle of systems 15
    3.1. Design 16
    3.2. Data 17
    3.3. Development 18
    3.4. Deployment 20
    3.5. Democratic pillars in relation to AI lifecycle 21
    3.6. Additional categorizations of AI governance 25
4. Building a culture of democratic AI governance 27
    4.1. Organisational foundations under the AI Act 27
    4.2. Important aspects of democratic AI governance 28
    4.3. Practical guidelines for implementation of high-risk AI systems 30
    4.4. Guidance for limited-risk AI systems 32
5. EU policy roadmap and recommendations 35
    5.1. Policy categories 35
    5.2. Roadmap 37
    5.3. Recommendations 44
6. Conclusion 49
References 50

List of Figures

Figure 1 – Pillars of democracy and how they are potentially affected by AI 10
Figure 2 – Illustration of how pillars of democracy intersect with AI lifecycle 15
Figure 3 – How different layers of the AI lifecycle relate to pillars of democracy, with the strongest connections marked with a + sign. 24
Figure 4 – Visualisation of the policy roadmap, detailing policy actions to be taken between 2026 and 2035 to advance democratic AI governance in Europe. 39

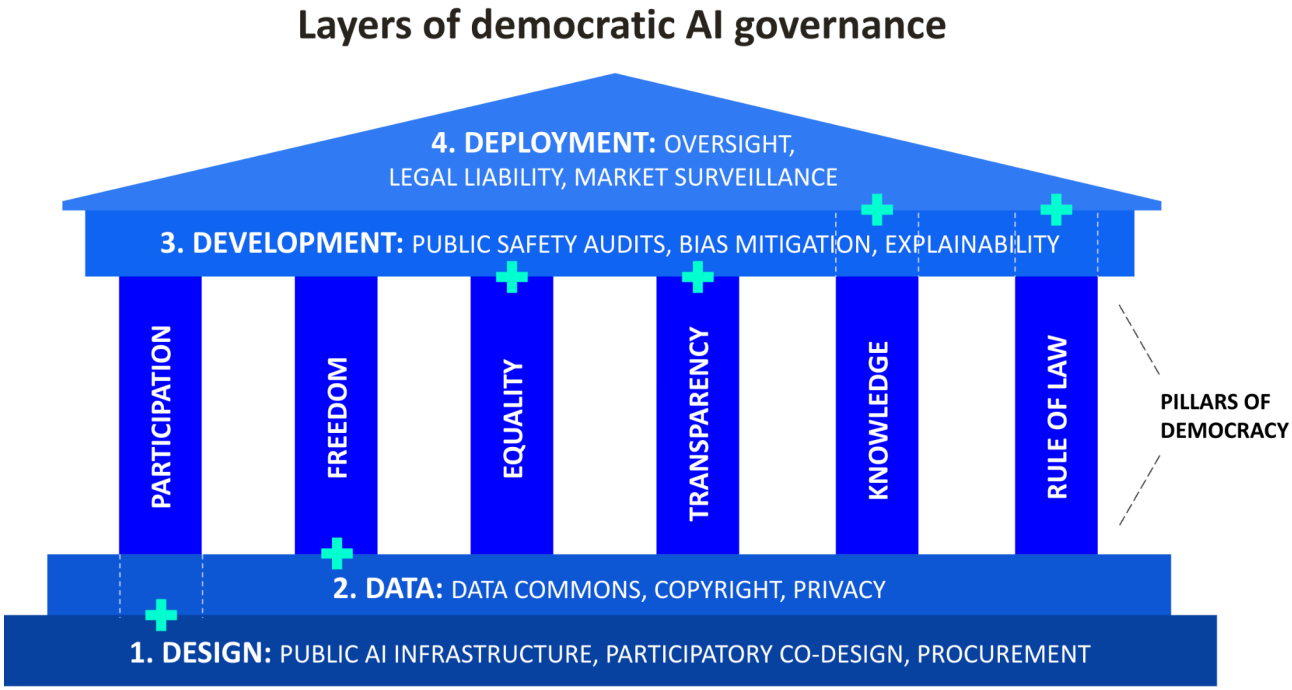
List of Tables

Table 1 – Most relevant stages of the AI lifecycle for each pillar of democracy 22
Table 2 – The overall policy roadmap divided into short-, medium- and long-term actions across five policy categories. 40

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

Executive Summary

The current AI landscape is characterised by rapid technological development and accumulation of power into the hands of private companies. This underscores not only the importance of regulation to keep up with technological progress, but also ensuring that AI governance upholds democratic principles. This report presents a holistic framework for democratic governance of AI and big data, which combines the pillars of democracy with the lifecycle of AI systems. The framework's key insight is that while democratic impacts are most prominent in the deployment stage of AI, such safeguards should cover the whole lifecycle of AI systems. Not doing so risks delegating democratically influential design, data and development choices purely to private AI companies. As such, democratic AI governance should be approached holistically, also covering the infrastructure underlying the technology.



To implement this framework, the report also includes policy recommendations and a roadmap, co-created in interaction with AI policy experts in KT4D use case 1. The roadmap consists of five policy categories, outlining steps toward democratic AI policy until 2035: regulatory enforcement, public AI infrastructure, investments and innovation, AI literacy as well as research and standards. Thereby the report seeks to guide EU policymakers in governing the disruptive nature of knowledge technologies and AI systems to reinforce democracy.

Reading guide: For an introduction into recent developments in European AI and digital policy and values underlying it, see section 1. If you are a researcher seeking a better conceptual understanding of how AI affects different pillars of democracy (e.g., equality, freedom), read section 2. Those interested in the technical governance mechanisms across the lifecycle of AI systems, please refer to section 3. For AI companies or developers seeking practical guidance on organisational culture and democratic compliance under the European AI Act, focus on section 4. Lastly, European policymakers and civil servants are encouraged to read section 5 on policy roadmap and recommendations for democratic AI governance.

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

Introduction

The rapid advances in AI systems and their widespread deployment in society creates urgent challenges that require policy and regulatory responses. The societal disruption, centralization of power, opacity, and the speed of change inherent to current AI advances all generate risks to democracy. These technologies create considerable tensions from economic, psychological and democratic perspectives, along with safety considerations (Bengio et al., 2025). Most importantly, from a democratic perspective, AI and big data threaten to significantly centralize power to a few large AI companies, given the prohibitive training requirements of AI models in terms of data and computational resources (Widder et al., 2023).

The EU's approach to AI governance is largely based on product legislation to ensure AI systems that enter EU markets are safe for consumer use. This raises the risk that the EU regulation fails to genuinely engage with the wider sociotechnical issues and cultural disruptions brought forward by AI systems. While AI can directly violate fundamental rights through the rise of algorithmic discrimination and personal data breaches,[1] there is a need for a larger infrastructural view if democracy is to be strengthened through the use of AI (Selbst et al., 2019; Kaltheuner et al., 2024). This means considering the entire lifecycle of AI systems and the infrastructure that underlies them, rather than only focusing on downstream deployment of systems. It remains challenging to realize democratic AI governance if the digital infrastructure that grounds the modern public sphere is owned by private platforms. Hence, there is a need for a holistic policy perspective that considers enforcement of AI regulation, investments into public AI infrastructure and greater digital literacy along other considerations.

The structure of this report is the following. First, we give background on the EU's technology and AI policy in relation to international initiatives to clarify how democratic values are reflected in these policies (see Bakiner, 2023; Lavorgna, 2024). Second, six fundamental pillars of democracy, such as transparency and the rule of law are outlined, covering both potential harms and benefits AI poses to them. Third, existing AI governance approaches (e.g. data, privacy, and compute governance) are mapped against these different pillars, to more fully encompass how democracy should be considered across the lifecycle of AI systems. Fourth, concrete organisational and legal practices for democratic AI governance under the AI Act are outlined. Fifth and last, we present a policy roadmap and recommendations for European policymakers to realize democratic AI governance from a holistic, infrastructural perspective.

[1] See, e.g., the Italian Data Protection Authority's press release of 29 November 2024 titled 'Garante privacy a Gedi: attenzione a vendere i dati personali contenuti nell'archivio del giornale a OpenAI perché li usi per addestrare gli algoritmi', available at https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/10076913, which highlights the difficulty in facilitating the right to object provided for under the GDPR, in the context of AI training and more generally, a lack of transparency in relation to data used for training AI.

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

1. Values behind European AI and technology regulation

The European Union has firmly established itself as a pioneering force in digital and artificial intelligence policy, a commitment clearly reflected in the previous European Commission's (2024a) priority 'A Europe fit for the digital age'. While the EU utilizes a broad range of policy instruments, in digital domain this has centered on regulation, put into motion through the 'big five' acts, encompassing the Data Act, Data Governance Act, Digital Markets Act, Digital Services Act and Artificial Intelligence Act (European Commission, 2024b). In particular, the AI Act stands as a testament to the EU's determination to set global standards, marking a departure from the more *laissez-faire* approaches of the US and the state-centric model of China. By leveraging the Brussels effect, the EU not only aims to foster innovation and trust within its internal digital market but positions itself as a global arbiter in the ethical deployment of artificial intelligence. The Brussels effect (Bradford, 2022) refers to the extraterritorial impact of EU's policies, as companies and countries outside the region often adopt its more stringent rules to access the EU's lucrative market. Yet spurred by the Draghi (2024) and Letta (2024) reports, the Commission is currently advancing an aggressive regulatory simplification in service of competitiveness, implemented through Omnibus packages. In the field of digital and artificial intelligence the normative weight previously given to the Brussels effect – alongside the belief that regulation can enable innovation – is being radically transformed by the introduction of the EU's 'simplification revolution' (European Commission, 2025a).

1.1. Development of the EU digital policy and legislation: from embedded ethics to a risk-based approach

Research on the role of ethics and democratic values in AI governance is emerging as the EU's digital policy is taking concrete shape. Paul (2022) identifies three key approaches to AI regulation in social science literature. These approaches can be useful in understanding the political dynamics behind AI regulation, although they also come with blind spots. First is **the normative project of applied ethics**, which focuses on generating ethical and legal principles for protecting individual rights. However, this perspective tends to overlook the discursive struggles and local interpretations of law and ethics. **The technocratic rational choice** emphasizes the balance between acceptable risks and economic benefits. While useful, it is limited by its tendency to prioritize regulators' self-interest and to disregard global and local differences in norms. Furthermore, the approach tends to bypass democratic decision-making by constructing an apolitical concept of regulation with limited room for values and democratization measures. Lastly, **the politico-economic project (or corporate capture theory)** examines regulation through the lens of power relations and ecosystems, framing it as a political project at the risk of being compromised by private interests. Yet on a closer look, this view seems to rely on a false dichotomy between states and businesses while neglecting national regulators' economic motivations and the principles of good governance that underpin regulatory decisions. Hence, there is no ultimately 'correct' way of interpreting the motives behind regulation, which calls for a nuanced understanding of different incentives.

At the level of technology policy, the role of ethics seems to have shifted and become more limited over the past two decades. In her discourse analysis on the use of ethics in the EU policy documents, Blagovesta (2024) describes this shift as one from good governance to better regulation, where ethics is increasingly seen as a

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

way to unclog the innovation process. Referring back to Paul's (2022) first approach (*the normative project of applied ethics*), one could also see the drafting of this framework as an example of this. As for an application of the *Technocratic rational choice endeavour* and *The politico-economic project* approach, the contributions of the experts in the Delphi survey (see section 5) included insights on weighing ethical issues against innovation and industry interests, which are also core interests for the EU. The context of this framework and KT4D project is specifically EU-focused, which carries different ethical emphasis to other regions. This has implications for the scope of the framework and may create tensions in some contexts. For instance, robust AI regulation is generally objected to by the big tech in the US but welcomed to a greater degree by EU start-ups that may benefit from a more "even playing field".

While ethics has become embedded in the EU digital policy in varying ways, the grounding of the EU digital governance as stated in the *Shaping Europe's digital future* -policy programme, has been to build up a European approach to digital technology, steered by the Declaration on Digital Rights and Principles (2022). The Declaration is said to "promote a digital transition shaped by European values" and explicitly builds on "primary EU law, in particular the Treaty on the European Union (TEU), the Treaty on the Functioning of the European Union (TFEU), the EU Charter of Fundamental Rights and the case-law of the Court of Justice of the European Union, as well as on secondary EU law" (Preamble). The opening statement of the Declaration states:

> "The digital principles included in the Declaration are intended as essential concepts, based on common European values, and serving as guidance for a human-centred, secure, inclusive, and open digital environment, where no one is left behind." (p1).

It therefore includes references to both key EU treaties and global governance frameworks. The document also explicitly states that these principles should be promoted in international forums beyond the EU member states. The Declaration puts forth a list of principles, which should be guiding policymaking, technology development and deployment, including "digital principles, to serve all Europeans, along the following lines in particular: putting people at the centre of the digital transformation; solidarity and inclusion; freedom of choice; participation in the digital public space; safety, security and empowerment; and sustainability." (p.4).

Guided by these principles, the development of EU's AI policy and regulation can be seen to have progressed towards robust legislation gradually over the years in roughly three phases. First, with the initial era of ethics and principles (2017-2019), the EU was developing approaches and tools for tackling emerging issues generated by the increased use and gathering of big data, quickly advancing AI development and the expanding deployment of algorithms across societies and political sectors. Building on this, the direction of policy and regulation was more geared towards more specific domains, with gradually developing policies and governance initiatives like the white paper on AI (2020-2021). This development reached its peak with the passing of the AI Act and other concrete policies and legislation, like the Digital Services Act and Data Act (2022-2024). This progress is also detectable in the gradual shift in frames of reference of the legislation. Whereas GDPR was dealing with fundamental rights and their application to practices in collecting, sharing and control of data, the recent European AI Act is based on a new legislative framework, such as digital markets and product legislation. However, since 2025 EU's AI policy has shifted drastically due to economic competitiveness and deregulatory pressures.

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

1.2. Shifting AI policy priorities

Regulatory approaches to digital and AI policy have markedly shifted as security, strategic autonomy and competitiveness have become organising principles for the EU since 2025. There is, first, a tendency to present AI policy in harmonious relation to other policy domains and democratic values, such as sustainability and security. Yet, the EU Digital Services Act and the Digital Markets Act neglect any mention of environmental or sustainability considerations, whilst the AI Act incorporates few, watered down, limits to AI's environmental impacts (Hacker, 2024). To the extent risks are identified, these regulations seek a corrective governance posture: seeking to limit risks without grappling with their systemic nature. Second, risk is being geographically displaced beyond EU borders. The Critical Raw Material Act governs the sourcing, extraction, supply of minerals central to AI value chains. Minerals may be designated 'critical' due to their strategic economic importance, the presence of supply chain risks, EU supply vulnerabilities, import dependencies and long-term security concerns about global "systemic rivals" (Leonelli, 2025). Designated 'Strategic Projects' largely override any concern of the effects of extraction on affected non-EU communities (Art. 7 and 10 CRMA; Robinson, 2023).

Third, the European Democracy Shield (EUDS) continues the EU's janus-faced approach to the relationship between democracy and digital and AI policy. It continues from previous policy initiatives that aim to de-risk systemic security threats that AI and digital technology may pose for democratic processes and institutions, such as disinformation. At the same time, it also looks to encourage the ways in which such technologies can bolster participation, communicative practices and civic engagement. The EUDS recommends the preparation of an incident and crisis protocol under the Digital Services Act to allow states to undertake rapid and coordinated responses to large scale transnational disinformation operations. It continues its non-obligatory approach in encouraging the uptake of its Code of Conduct on Disinformation as well as in its guidance on "fair, transparent, human-centered and responsible use of AI in electoral processes" (European Commission, 2025b). The newly formed European Centre for Democratic Resilience is another enabling body tasked with cooperation, capacity building and minimal coordination between Member States. On the other hand, the EUDS encourages the use of digital and AI technologies for building civic engagement. Digital and AI literacy is conceived as central to democratic resilience, with both funding and educational recommendations central to emerging EU policy (ibid. 21). Stimulating innovation in civic tech enables the Commission's agenda to broaden and deepen democratic participation (ibid. 24). Overall, the EUDS takes a largely epistemic approach to both democratic participation as well as AI and digital technologies: the risk of the latter and the potential of the former lies in the quality of informational input.

Fourth, Europe's AI industrial policy strategy seems to displace the EU's previous emphasis on mitigating the risks of scaling the use of technology. Rather than relying on strong regulation-induced conditionalities, the EU Continent Action Plan and the Apply AI Strategy aim to mobilise innovation and kick-start the EU's industrial policy in digital and AI fields through a de-risking approach that mobilises private capital on the basis of public-private partnerships and the loosening of state aid frameworks (see the Clean Industry State Aid Framework; European Commission, 2025c). Fifth and finally, the Commission's simplification agenda challenges both risk-mitigation and rights-based regulations. Digital and AI regulations are being watered down through Omnibus processes, whilst the Commission has seemingly adopted a policy stance it had previously rejected: the false dichotomy between regulation and innovation (Bradford, 2024). Simplification is framed as a core driver of innovation and European industrial policy strategies (European Commission,

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

2025d). In late 2025, the Commission produced the Digital Omnibus (2025e) and Digital Omnibus on AI (2025f) that redefined the scope of previous digital rights, watered down risk-mitigation accountability mechanisms for high-risk AI systems and ensured that AI literacy worker training was no longer mandatory for companies. The simplification agenda also, more fundamentally, challenges the ethics-centered approach seen in better regulation discourses. Better Regulation requirements are both being circumvented by justifications of urgency in Omnibus packages, as well as being made subject to simplification processes themselves (ibid.).

Alongside the EU-driven governance initiatives and regulation, the most powerful players in digital geoeconomy, namely the United States and China, have followed their own choices and pathways in developing legislation. Commentators and researchers have closely followed these developments in recent years (see Donoghue et al., 2024). As a summary of the differences between these three examples, Chun et al (2024, 26) argue that the EU approach has the strengths of being "a coherent, universal, risk-based regulatory framework with strict and well-defined penalties", while it is also being seen as lacking in fostering innovation and suffering from lack of anticipatory capacity to the possible challenges in implementation in concrete AI use cases. They refer to China as a synthesising case between the US approach, where the use-case specific laws are at the heart of the legislation, along with general guidelines. The outcome of this is a centralised framework, with components for registration, testing and monitoring. However, the Chinese model also includes both direct and indirect initiatives for supporting innovation and economic growth (ibid.). The US, in turn, has counted on a strongly market-driven approach to legislation, which leans on self-regulation and stakeholder competition. The approach has been criticised for counting too much on the adoption of soft measures. Recent examples from the US also indicate that some of the state-level initiatives (in California notably) could challenge the Federal approach and possibly drive a stricter way forward in legislation.

1.3. Beyond the EU: examples of legislative initiatives and guidelines

While the EU has been prominent in developing legislation, significant initiatives have also been introduced internationally. In fact, some have argued that there has been a shift from 'race to AI' to a 'race to AI regulation' globally (Smuha, 2021). The multilateral **global governance** initiatives have come to include several initiatives. UNESCO (2021) has put forth recommendations for ethics and AI, accompanied by a Global Ethics and AI Governance Observatory. OECD's Principles for Trustworthy AI were adopted already in 2019, including both value-based principles and recommendations for policymakers. The Global Partnership for AI (GPAI) has promoted a joint approach to AI governance by addressing dimensions like the future of work, responsible AI and data governance, having now entered a partnership with OECD. In late 2023, safety-oriented agreements such as the G7 Hiroshima agreement and Declaration of the Bletchley Park AI Safety Summit were announced. Since the AI Seoul Summit in 2024, the collaboration between nationally established AI safety institutes (Australia, Canada, EU, France, Japan, Kenya, the Republic of Korea, Singapore, UK, US) has strengthened into an international network of AI Safety Institutes, which met for the first time in November 2024 (European Commission, 2024c). Moreover, The Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law was opened for signatures in September 2024 as the first international legally binding treaty in the field (CoE, 2024).

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

After the initial safety focus of the Bletchley Park and Seoul AI Summits, these international forums have shifted more towards innovation. While the AI Action Summit hosted by France in February 2025 gathered representatives from over 100 countries, the final declaration on inclusive and sustainable AI was signed by only 61 nations, with key players like the US and UK opting out of specific binding commitments to prioritize innovation and national competitiveness. The next AI Summit in February 2026 takes place in New Delhi, India, focusing on impacts of AI. Meanwhile, the United Nations has moved forward with its Global Digital Compact implementation by establishing the International Scientific Panel on AI to provide independent scientific evidence for AI policymaking as well as the annual Global Dialogue on AI Governance as an inclusive platform for stakeholder discussion, in 2025. Nevertheless, concerns remain about the capacity of the slow, consensus-based UN system in responding to the pace of technological development.

Outside the EU, there have also been other notable **national level developments**. President Donald Trump's Executive Order on Removing Barriers to American Leadership in AI in January (The White House 2025a), America's AI Action Plan in July (2025b) and Ensuring a National Policy Framework for AI in December (2025c) explicitly revoked the Biden administration's 2023 Executive Order (14110), effectively dismantling the nascent safety reporting requirements for frontier models in favour of a deregulation focused on infrastructure and energy. Consequently, the NIST AI Risk Management Framework (2023) remains voluntary in the broader private sector, and the enforceability of federal safety reporting and risk-management actions envisioned under the earlier executive order has been rolled back. Similarly, the UK has reaffirmed its "light-touch" regulatory stance. Despite the reintroduction of the Artificial Intelligence (Regulation) Bill in March 2025, the UK government largely maintained its pro-innovation approach, formalizing a partnership with the US in February 2025 to prioritize economic growth over statutory restrictions. The full text for the Technology Prosperity Deal between the countries was released in September 2025 (UK Prime Minister's Office, 2025). In contrast, other regions have faced legislative stalls; in Canada, the long-awaited Artificial Intelligence and Data Act (AIDA), part of Bill C-27, effectively died on the order paper in January 2025 due to the prorogation of Parliament, leaving the country without its anticipated comprehensive framework. Meanwhile, China adjusted its strategy by removing its comprehensive AI Law from the 2025 legislative agenda, opting instead for targeted measures. This was exemplified by the mandatory labeling rules for AI-generated content that were issued in March 2025 and came into force later that September, designed to maintain social stability while fostering industrial growth.

A third form of initiatives are the ones generated at **the industry level**, including the Partnership on AI and the EU Artificial Intelligence Pact, which reached its first hundred signatories by companies in September 2024 (European Commission, 2024d). These have been accompanied by voluntary commitments by leading AI companies and examples of responsible scaling policies. For instance, the EU's Code of Practice on General-Purpose AI (GPAI) is a voluntary instrument that helps the providers of GPAI models to comply with the AI Act's legal obligations. The final Code of Practice was released in July 2025, composed of transparency, copyright and safety and security chapters – the last of which only applies to the most advanced models deemed to pose systemic risks (European Commission, 2025g). The above-mentioned CoE treaty on AI is also currently being implemented in an attempt to facilitate the interests from different sides. While the treaty opened for signatures in late 2024, its application to private companies remained a point of contention throughout 2025. Major non-EU signatories, including the US and UK, utilized the treaty's flexibility clauses throughout 2025 to exempt their domestic private sectors from direct treaty obligations, seeking to protect their industries from what they viewed as European-style regulatory overreach (Rotenberg, 2025).

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

While current global initiatives focus heavily on the tension between safety and innovation, they often overlook the democratic values that underpin regulatory legitimacy. The shift toward deregulation and industry exemptions suggests that technical frameworks are being prioritized over the protection of civic values and fundamental rights. From the perspective of democratic AI governance, it is necessary to look beyond technical checklists and examine how such policy frameworks uphold or erode the core pillars of democracy. The following section introduces a framework designed to analyse how AI impacts these democratic foundations.

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

2. Pillars of democracy

Democracy, in its various forms, is anchored by fundamental principles that ensure the legitimacy, stability, and adaptability of governance. The precise elements that define democracy have been the subject of extensive scholarly debate, but key dimensions — such as participation, freedom, equality, rule of law, knowledge, and transparency — consistently emerge as core features. These pillars resonate with the frameworks laid out in seminal works on democracy, including Robert Dahl's polyarchy model (1971), which emphasizes effective participation and enlightened understanding as well as David Held's typology of democratic models (2006), which explores democracy's historical and ideological evolution. Such pillars also draw from T. H. Marshall's theory of citizenship (1950), which categorizes civil, political, and social rights as essential to democratic life. Aside from the above-mentioned traditional 'models of democracy', this also aligns with Mark Warren's problem-based approach to democratic theory (2017), which emphasizes empowered inclusion, collective agenda and will formation, as well as capacity for collective decision-making.

The progress in AI presents both opportunities and challenges for the pillars of democracy. AI technologies influence democratic processes and structures in novel ways, raising critical questions about how they intersect with and potentially reshape core democratic values. For instance, AI can enhance participation through tailored engagement platforms, but it may also undermine it through algorithmic manipulation or misinformation. Similarly, AI's role in enabling personalization and predictive analytics challenges traditional notions of freedom, as surveillance systems risk infringing on privacy and autonomy. Existing literature on the intersection of democracy and technology provides insights into these tensions. Scholars like Daniel Kreiss (2016) and Philip Howard (2010) examine how digital tools influence political participation and deliberation, while thinkers such as Shoshana Zuboff (2019) criticize the implications of data-driven governance for individual freedoms. In the realm of equality, Kate Crawford (2021) and Safiya Umoja Noble (2018) highlight the structural biases encoded in AI systems and their implications for marginalized communities. Moreover, Lazar (2024, 2025) has highlighted how AI causes considerable shifts in macro-level power dynamics and disparities, whilst reshaping social norms and institutions. This concentration of power can erode democratic governance, as reflected in recent antimonopoly and antitrust governance practices (see Khan, 2025).

2.1. Impact of AI on democratic pillars

We adopt a six-pillar framework — participation, freedom, equality, rule of law, knowledge, and transparency — to analyse AI's democratic impact, drawing on a rich body of academic literature. This typology, while not exhaustive, captures the multifaceted nature of democracy and offers a robust lens for understanding how AI technologies influence its core values. By examining AI's effects across these pillars, we aim to contribute to the growing discourse on how democratic principles can be safeguarded and strengthened in an era of rapid technological change. Figure 1 summarizes these relationships before describing them in detail.

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

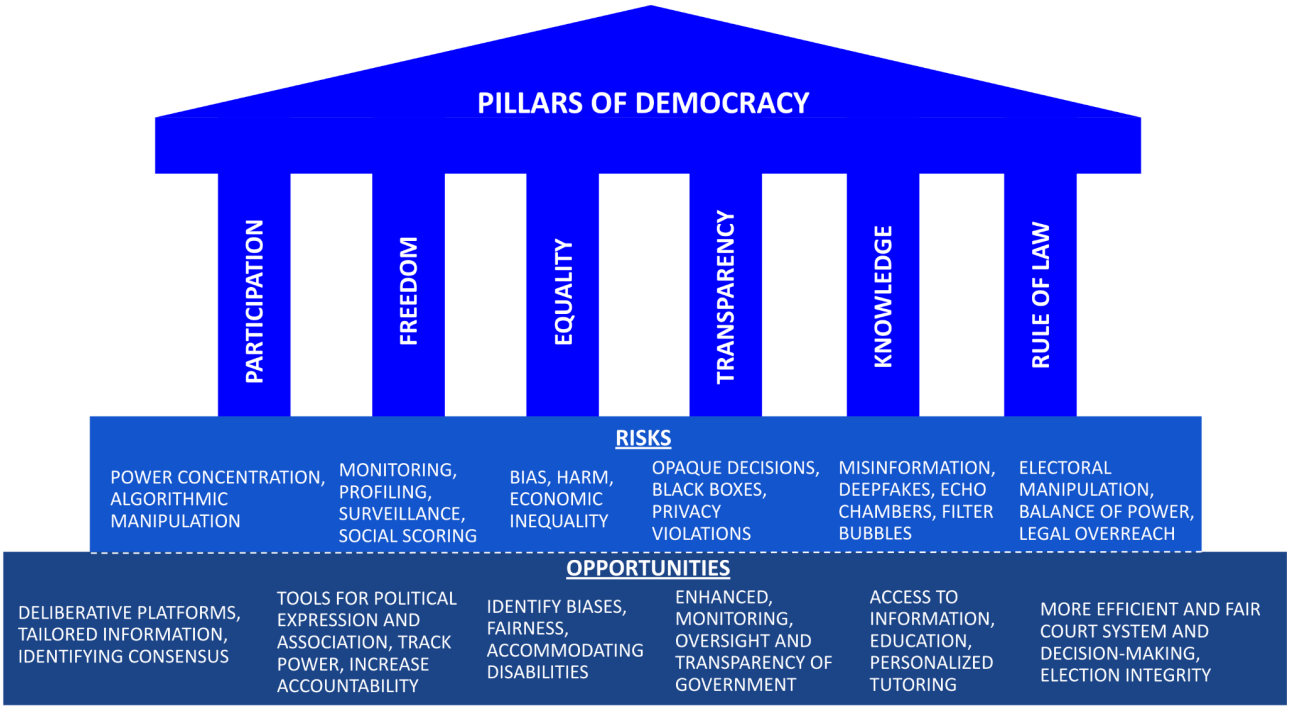


Figure 1 – Pillars of democracy and how they are potentially affected by AI.

1. Participation

A democratic system rests on active political participation of citizens in the political process and decision-making. The authority of the government is created and sustained by the ongoing consent of its people, either through more direct means or elected representatives. Hence, the government should be responsive to the will of the citizens, ensure all citizens have a right to participate and influence policy, as well as take steps to allow citizens to properly realise this right. Participatory elements for legitimacy of democratic decision-making are especially stressed by theories of participatory and deliberative democracy (Bächtiger et al., 2018).

AI can affect political participation for the better or worse:

- Algorithms can be used to influence and manipulate public opinion, especially on social media platforms by shaping the information citizens see. For example, AI-generated misinformation has been used to influence elections (Shukla & Tripathi, 2024). Moreover, there is a more general risk of concentrating power to those who design, develop and deploy – and now have access to – AI systems, limiting the effectiveness of public participation in decision-making.

- AI can also enhance civic participation and accessibility by creating more representative channels and platforms for citizens to deliberate and impact decision-making. Moreover, AI systems can make policy information more understandable to citizens (e.g. translation, visualization, tailored information), help them express themselves politically, facilitate deliberation and identify consensus to decision-maker (Combaz et al., 2024). AI's incorporation into civic tech also has the potential to scale democratic innovations, increasing their quality and number of participants (McKinney, 2024). Yet, better deliberative tools do not automatically equal more legitimacy, as civic tech often reshapes the very public it purports to serve (Hahn & Farrell, 2025).

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

2. Freedom

Democracy requires political freedom and civil liberties for its citizens, including the freedoms of association, assembly, and expression. In a more traditional sense, political freedom includes negative freedom from external oppression or coercion by others, as well as more positive freedom to exercise one's rights and capacities (Berlin 1958). Aside from traditional theories of liberal democracy, neorepublicanism defines freedom as non-domination, the independence from arbitrary power (Pettit 1997, Skinner 2025).

AI can influence the freedom for political action or speech by:

- Enabling extensive monitoring, profiling and surveillance, which can deter free expression and assembly of citizens. These issues are especially relevant to facial recognition and social scoring systems in context such as policing and border control (Ashraf, 2020). More advanced AI models could also potentially violate freedom by evading control and being misaligned with human interest (Dung, 2023).
- On the other hand, AI systems can facilitate freedom of expression and association by aiding individuals to better express themselves with likeminded people (Helberger et al. 2020). Moreover, AI tools could also be used to track and monitor the state's use of power to hold them more accountable.

3. Equality

Political equality of citizens is essential to a functioning democracy. This includes the protection of fundamental rights for all individuals as well as respect for diversity of backgrounds and political beliefs. Aside from egalitarians like John Rawls (1971) and G. A. Cohen (2009), social democrats tend to link economic redistribution and alleviation of structural inequalities with democratic equality. Such scholars argue that democracy requires a sufficient material and economic equity to maintain societal trust and a sense of political equality between citizens.

AI might impact social equality and power relations by:

- Inadvertently reinforcing and exacerbating existing societal biases and discrimination, thereby leading to underrepresentation and exclusion of certain groups in decision-making. Algorithmic bias can creep into the system through e.g. unrepresentative training data, thereby undermining equality (Kordzadeh, 2021). AI may also increase economic inequality by replacing cognitively demanding labour and thereby increase unemployment, given that AI development is predominantly driven by profit motives of private technology companies (Du, 2024).
- Promoting substantive equality (e.g., by identifying existing biases or choosing the right fairness metrics) and enhancing more transparent decision-making. Moreover, AI tools could be used to accommodate disabled people, and amplify other underrepresented views. It could also aid more equal economic outcomes and distribution of resources by allowing for, e.g., basic income or windfall clauses (see O'Keefe, 2020).

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

4. Knowledge

Democratic decision-making processes rely on a well-informed electorate and information ecosystem. Moreover, an independent press and media is arguably essential to democracy, serving as a "watchdog" that keeps citizens informed and helps to hold those in power accountable. Overall, access to information is essential for informed participation in political life. Theorists of epistemic democracy such as David Estlund (2008) stress the importance of knowledge, as their justification for democratic system is derived from its epistemic ability to make good and just political decisions.

AI influences the knowledge base of democratic societies through:

- Increasing the amount of misinformation and disinformation such as "deepfakes" that potentially distort people's view of reality, diminish trust in any kinds of information outlets and diminish the epistemic agency of citizens (Coeckelbergh, 2023). Moreover, algorithms can contribute to echo chambers and filter bubbles that polarize the political discourse.
- Enhancing access to information and improving educational outcomes through personalized tutoring and other means (Bilad et al. 2023). AI could also be used to reconcile differing views and identify consensus between parties.

5. Transparency

Transparency and accountability are essential values of democracy. Government's actions should be open to scrutiny and public officials held accountable for their decisions. Transparency is inherently connected to the justification for democracy – the openness enables citizens to trust the democratic process. Ideals of transparency are not only upheld by scholars like Archon Fung (et al. 2007) but also by movements such as open government, freedom of Information and open-source government, especially in the digital realm. In contrast, algorithmic decision-making is often criticized for its lack of transparency.

As such, AI can potentially affect transparency by:

- Leading to opaque decision-making in public services as more decisions are delegated to AI systems. "Black-box" algorithms that are not understandable or explainable to citizens make it challenging for citizens to hold the government accountable and challenge decisions if needed, especially if individuals are unaware that AI systems are being used (de Fine Licht, 2020). Moreover, the development of AI systems also risk privacy and data protection violations and entail exploitative data collection practices, often unbeknownst to individuals.
- Enabling enhanced monitoring, oversight and transparency of government activities and identifying possible corruption by making data available and understandable for citizens. Even so, these AI systems themselves should remain transparent (Köbis et al. 2022).

6. Rule of law

Funded by the European Union

KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

Rule of law is a backbone of democracy: individuals and institutions should be accountable to laws that are applied and enforced equally to all, with mechanisms in place to prevent abuses of power. Elections should be conducted in a free and fair manner without voter suppression. Moreover, democratic decision-making rests on separation of powers between different branches of government (traditionally executive, legislative, and judicial) that hold each other accountable through a system of checks and balances. These views are espoused by legal positivists or on the other hand constitutionalists such as Ronald Dworkin (1986).

AI can have an impact on the rule of law through multiple mechanisms:

- AI systems can be used for electoral manipulation to weaken integrity of elections (Shukla & Tripathi, 2024). Unchecked use of AI by any single branch could upset the balance of power if it centralizes decision-making power. Moreover, the use of AI systems can amount to overreach and violation of individuals' rights under the pretext of legal enforcement.
- AI could help in legal decision-making, providing a both more efficient but also transparently and consistently operating court system. Moreover, it could improve transparency within government branches, analysing data for signs of overreach (Köbis et al., 2022).

2.2. Intersection of the pillars in relation to AI

It is worth noting that these pillars of democracy are interlinked and cross-cutting, and there are multiple scholarly typologies about fundamentals of democracy. For example, Habermas (1996) integrates knowledge, participation, and equality in his framework of deliberative democracy. Multiple categorizations or typologies of the risks that AI poses for democracy can be conceived. One such categorization distinguishes risks in terms of automated decision-making, democratic participation, and electoral processes. One can first think of risks of automated decision-making (ADM), where the problems relate to algorithmic discrimination, lack of transparency and erosion of human autonomy. Second, there are the risks for democratic participation, relating to the erosion of the public sphere, concentration of power to AI companies as well as economic inequality through unemployment. Third, from an electoral perspective, AI can be misused for voter suppression, disinformation and surveillance. Particularly challenging are the issues around open-source AI models. While openly available AI models are important for increasing transparency around AI, this accessibility has also been argued to increase the risks of misuse, like AI-driven misinformation (Seger et al. 2023a). Democratic governments therefore face a safety versus openness trade-off where safeguarding democracy might require limiting access to advanced AI models, thereby concentrating power to a few AI companies.

Other related typologies of AI's impacts on democracy have been offered. For instance, Andreas Jungherr (2023) identifies the domains of democracy affected by AI as individual self-rule, equality between citizens, election integrity and autocratic competition between nation states. The self-rule is the most multifaceted of these, covering how AI shapes information environments, economics of news, speech and manipulation. These issues include opaque algorithmic platforms that affect the political information citizens are exposed to, AI-based moderation of political speech based on behavioural predictions, automated generation of news content through AI and incentives for distributing (mis)information. Moreover, self-rule is also potentially

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

undermined through the increased power of private AI companies and by strengthening expert rule through better data-driven predictions (Jungherr 2023, 6). More generally, AI-driven filtering, recommendations and content is likely to strengthen the intermediary structures of the public sphere, hide challenges to the political status quo and fortify control by gatekeepers (Jungherr & Schroeder 2023).

Mark Coeckelbergh (2022) focuses on how AI technologies can undermine democratic values and practices through mechanisms such as bias, discrimination and the creation of echo chambers, which may lead to what he terms "machine totalitarianism." He also emphasizes the role of AI in enabling pervasive surveillance and promoting self-discipline through data-driven processes, while highlighting broader implications for non-human agency and environmental politics. In contrast, Noorman & Swierstra (2023), drawing on Warren's (2017) problem-based approach, categorize AI's effects based on its interaction with key democratic practices such as recognizing rights, resisting power, deliberating publicly, voting, representing, joining collectives, and exiting relationships.

The pillar approach to AI's effects on democracy is instructive because it goes beyond the easy-to-identify direct harms of AI, like intentional misuse, whether that is disinformation, electoral manipulation or surveillance. The indirect structural effects, such as centralisation of power to few private companies across the AI value chain, lack of democratic oversight and diminishing epistemic agency of citizens due to increasing reliance on AI, tend to remain unnoticed. These more mundane and invisible effects might pose problems that are as – if not more – important for democracy as misuse or specific high-risk systems. These systemic disruptions on democratic culture are arguably challenging for regulations such as the AI Act to address, as they rely on a risk-based approach to regulating specific systems.

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

3. AI governance across the lifecycle of systems

In light of the above examination of how AI can impact different pillars of democracy, this section outlines the different AI governance approaches spanning the lifecycle of AI systems: design, data, development and deployment. Identifying different intervention points in AI governance along the AI lifecycle (such as privacy and compute governance) can be informative, since it sheds light on how well risks to democracy are reflected in current AI regulatory frameworks. The threats posed to each of the above pillars of democracy require different governance approaches to be tackled. Mapping these governance approaches helps to identify which temporal points along the lifecycle of AI systems are most important and opportune to intervene in, when it comes to different democratic harms. Figure 2 presents a simplified overview of which democratic pillars are the most relevant for each part of the AI lifecycle (see section 3.5).

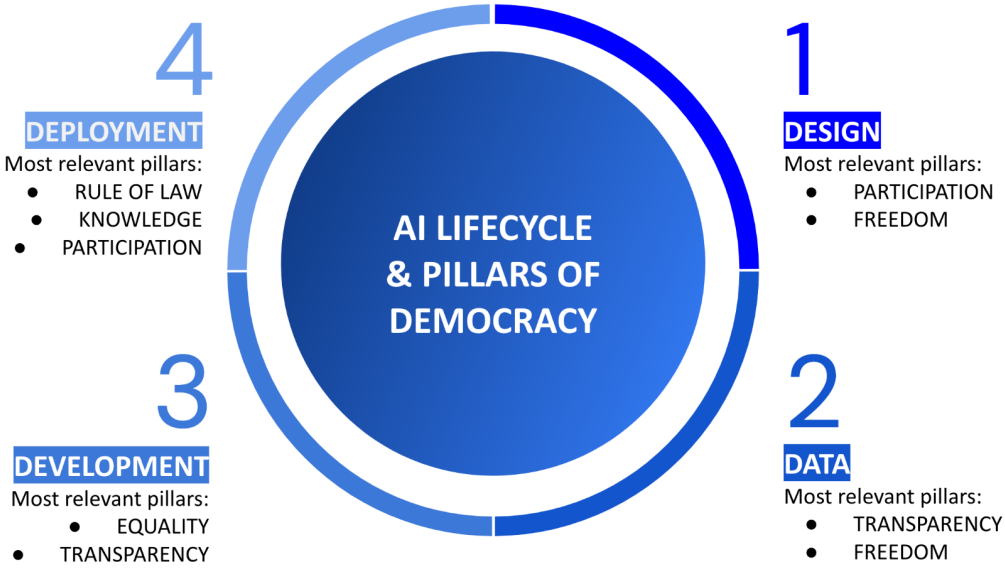


Figure 2 – Illustration of how pillars of democracy intersect with AI lifecycle.

To better capture what different regulatory interventions target along the AI lifecycle, below is a standard depiction of an AI lifecycle:

- Design
  - Pre-design and procurement
  - Compute governance & licensing
  - Digital public infrastructure
  - Participatory design requirements
- Data
  - Data quality criteria
  - Copyright legislation
  - Privacy governance
- Development

KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

- Safety & risk assessment standards
- Bias mitigation
- Explainability and transparency policies
- Model audits and evaluation

- Deployment
  - Staged release and access policies
  - Monitoring and oversight
  - Liability and legal remedies
  - Competition and market legislation

To summarize, the lifecycle of AI systems can be presented as consisting of four stages: 1) Design and requirements setting, 2) Training data collection and processing, 3) Model development and training, and 4) Deployment of the system. As outlined, different policy and governance domains are relevant depending on the lifecycle stage in question. These stages are discussed below in relation to the risks and opportunities they pose in relation to the pillars of democracy. The focus here is on policies and governance interventions enacted by governments rather than purely voluntary commitments by private AI companies. That being said, there are obviously different degrees of regulatory stringency – ranging from soft law such as incentives to hard law prohibitions of certain practices.

3.1. Design

In the first phase of the AI lifecycle, design, the emphasis is placed on the process and design requirements for the system as well as the physical infrastructure on which it is being built on. This includes defining the AI system's objectives, rationale for use, its necessity and potential impacts, taking into consideration the societal context where the system is planned to be deployed. Moreover, it sets requirements for the development stage, such as performance, accessibility and maintenance of the system. This process of designing and planning the system is arguably one of the most crucial steps from the viewpoint of equality and democracy. This is even more so in the context of public sector procurement, where agencies lack oversight of the AI development process. Moreover, requiring the participation of a variety of social stakeholders and citizens is important from the perspective of democratic equality. The key governance mechanisms that fall under design are expanded upon below.

Pre-design and procurement: This governance aspect focuses on the early stages of AI design, where the goals, necessity, and social context of the system are defined. It involves a clear articulation of the system's purpose, problem formulation, target users, and affected stakeholders, ensuring the system is designed to meet its intended objectives. Governance at this stage can also evaluate sector-specific risks and ethical considerations, considering alignment with democratic values from the outset. Setting these requirements is the first intervention point for AI regulation, which enables governments to influence how and for what purposes systems are developed, especially in public contexts. This is particularly the case when the public sector agencies themselves are procuring the system (Ojanen et al., 2022). Certain systems or outputs can also be prohibited from the start, specifically in public sector decision-making, e.g. through pre-development assessments.

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

Digital public infrastructure: Refers to publicly accessible digital resources that are essential to AI system development. Governance of digital public infrastructure focuses on creating AI resources designed to serve public interests, such as open datasets, public algorithms, or shared governance platforms. It can also cover data sharing that can serve as building blocks for a variety of applications. The common feature of all items of digital public infrastructure is that they can be integrated into many different systems and applications. These infrastructures aim to decentralize control, enhance transparency, and make AI development more collective, reducing the risks of monopolistic or authoritarian control by private AI companies (Eaves et al. 2024). Digital public infrastructure can be a tool for enhancing participation of citizens and accessibility of AI development, but also carries risks, such as locking-in societies into uncompetitive and costly infrastructure.

Compute governance & licensing: Compute governance refers to the regulation of the use of physical computational and hardware resources, such as GPUs and TPUs, which are crucial bottle necks in the development of advanced AI systems. Sastry et al. (2024) argue that computational resources, i.e., compute is especially amenable to governance because of the physical nature of AI chips and their market concentration, making them easily quantifiable, tracked and restricted. For instance, computational resources can be regulated in terms of ownership, access, geographic location and environmental impact of data centers. Compute governance covers both chip distribution (i.e., cross-border flow of AI-specific chips), and compute provider oversight (which requires AI compute providers to report usage above predetermined thresholds; Dennis et al. 2024). Moreover, compute restrictions can be coupled with licensing requirements, requiring AI developers to disclose, preregister and acquire government approval before large-scale training runs to use these computational resources. Licensing requirements can also apply to the overall development of systems, especially in compute-intense and high-risk settings.

Participatory design requirements: One governance leverage is to require the inclusion of diverse stakeholders, particularly underrepresented and marginalized groups, in the design of AI systems (Parthasarathy et al. 2024). Such participatory requirements aim to ensure that the systems developed address a broad range of societal needs and concerns. In addition, accessibility requirements may be imposed to ensure equal opportunities to use and access the service. This involves embedding user perspectives into the basic design of the system, such as the needs of minorities, and reasonable accommodations for people with disabilities designing the system, in order to reduce the risks of exclusion and increase trust in AI.

3.2. Data

Data governance addresses the processes and policies governing the collection, preparation, and use of data in training of AI systems, which is one of the central building blocks of AI systems alongside compute resources and algorithms. It can be defined as the practices, institutions and other rules on the management, sourcing, and quality of data used to train AI models (Janssen et al., 2020). As training data that reflects past inequalities is one of the main causes of algorithmic discrimination, this stage involves interventions that mitigate risks associated with biased, low-quality, or unethically sourced data. It also seeks to ensure that datasets align with regulatory standards in terms of copyright protection and individual privacy and data protection rules.

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

Data quality criteria: Data quality provisions establish standards and processes for origins, selecting, labeling, and preparing datasets to ensure their quality, thereby addressing one of the primary causes of algorithmic discrimination: unrepresentative or mislabeled data (Schwabe et al., 2024). These standards establish rigorous processes for data collection, curation, and validation, involving dataset auditing to identify biases, correcting inaccuracies and filtering out harmful content. This can involve requirements like datasheets, which document how data is collected, processed, and used. For instance, Article 10 of the EU AI Act (Regulation 2024/1689) specifies quality criteria for datasets in high-risk AI systems, ensuring they are both reliable and representative. Aside from purely technical interventions, such practices should also consider societal context and intended purpose of the system. Standards might also prohibit reliance on improperly obtained or overly sensitive data and discourage unethical practices, such as outsourcing annotation to precarious minimum wage workers.

Copyright legislation: Copyright governance addresses the use of protected intellectual property in datasets for training AI systems. Mechanisms like licensing agreements, public datasets or exclusion of copyrighted materials without permission aim to ensure that AI development does not exploit creative workers or disrupt business practices. Such debates have gained prominence with general-purpose AI models that are generally trained on large swathes of the internet like Common Crawl (Desai & Riedl, 2024). Improperly acquired datasets not only violate copyrights but may also contain democratically problematic content such as hate speech, violence and sexual imagery. Ensuring that training data of AI systems respects copyright laws is therefore an urgent priority under the current AI paradigm, which might require revisiting and strengthening the legislation in response to AI-based scraping of data.

Privacy governance: Focuses on safeguarding individuals' right to anonymity, confidentiality and protections for personal data in training of AI systems, ensuring compliance with data protection regulations like GDPR (Bennett & Raab, 2020). Privacy governance regulates how personal data is collected, processed, and stored, emphasizing informed consent and limiting intrusive surveillance. Such practices seek to prevent personally identifiable information appearing in AI training datasets and limit the risk of AI models like LLMs memorizing and leaking personal data, such as names, phone numbers, or email addresses (Ippolito et al., 2023). These risks can lead to privacy and data protection violations, identity theft, and broader security threats, emphasizing the role of strong privacy protections. Methods for mitigating these issues include pseudonymization, anonymization, and privacy-enhancing technologies, along with robust cybersecurity measures to prevent data breaches and protect proprietary model information, such as model weights. Given the potential for identity leaks, reconstruction of classified data and cyberattacks, data privacy becomes interlinked with security.

3.3. Development

The development phase of the AI lifecycle involves the model training, testing and validation to ensure that it is robust, fair, and aligned with societal expectations. This stage requires technical and procedural safeguards to mitigate risks and address the potential limitations and risks of the model, such as discriminatory biases or unreliable outputs. This covers the theoretical underpinnings and architecture of the model along with its learned parameters. Based on these evaluations, the model may need to be revised and retrained to improve performance and reduce unintended harms. Key components of model governance

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

include implementing bias mitigation methods to address harms, ensuring outputs are interpretable, conducting model evaluations to assess an AI model's capabilities and establishing robust risk management frameworks. These issues also benefit from standardized safety practices that are scalable and reliable across systems.

Safety & risk assessment standards: Safety standards establish protocols to identify, mitigate, and manage risks both before and after model training (Anderljung et al., 2023). Pre-training measures focus on maintaining model accuracy and adversarial robustness (e.g., adversarial training to prevent attacks like jailbreaking). Post-training procedures address emergent behaviours and potential misuse scenarios. Comprehensive risk assessments are crucial, evaluating the model's forecasted capabilities, societal impacts, and potential for dangerous misuse. AI developers can be obliged to plan and report measures to mitigate risks, ranging from security vulnerabilities to ethical concerns. Robustness testing ensures models sustain performance under adversarial or unexpected conditions. By implementing governance mechanisms to continuously monitor model training and assess reasonably foreseeable harms, organisations can minimize the chances of catastrophic failures or malicious use.

Bias mitigation: Involves implementing processes and tools to identify and reduce discriminatory algorithmic patterns embedded in the AI model, so as not to perpetuate systemic inequities. This includes technical interventions like adversarial debiasing or re-weighting datasets and institutional mechanisms like fairness audits (Gray, 2023). Governance frameworks can mandate bias assessment and mitigation across different parts of model training, e.g. pre-, in-, and post-processing phases. For example, pre-processing can involve balancing training datasets to improve representativeness, while post-processing can balance biased model outputs. These methods are important for ensuring AI systems provide fair and equitable outcomes, particularly in high-risk applications like hiring or criminal justice. However, there might be trade-offs between accuracy and fairness of the model that require additional guidance by regulatory agencies.

Explainability and transparency policies: Governance frameworks for transparency mandate that AI models provide interpretable outputs and clear explanations for their decisions. This is especially critical for high-risk applications like healthcare, insurance, finance, and judicial decision-making. However, the complexity of large AI models, particularly those with billions of parameters, often create an opaque 'black box problem' (von Eschenbach, 2021). Explainability towards citizens seeks to ensure democratic oversight by making the assumptions and logic of AI systems clear. Unexplainability poses largest risks when models generate harmful or biased outputs, such as discriminatory content or guidance for malicious activities. Governance frameworks can promote such transparency and explainability by requiring systems to be designed for interpretability, either through simplified architectures or post-hoc tools (e.g. causal and counterfactual explanations). These measures enable regulators, developers, and end-users to scrutinize AI decision-making processes and ensure accountability.

Model audits and evaluation: Audits and external evaluations of models provide impartial oversight to ensure AI systems meet safety, performance, and ethical standards. These evaluations, conducted at various stages of development, identify risks and failures that internal developer teams might overlook. General-purpose AI models can be assessed at least in terms of their a) general capabilities and limitations and b) societal impact and downstream risks. Key evaluation areas include general model capabilities, societal

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

impacts, and potential misuse risks (Shevlane et al., 2023). Audits assess how design choices (e.g., architecture, training data, parameters) influence model performance in controlled or real-world scenarios. They may also examine downstream effects, such as amplification of biases, security vulnerabilities, or environmental impacts. External audits may involve independent third parties or government bodies with controlled access to fine-tuning APIs or other technical tools for 'red-teaming' the model. Impartial and regular audits enhance regulatory compliance, helping to ensure that unsafe models are not deployed.

3.4. Deployment

Deployment is a critical phase in the AI lifecycle, where models are integrated into and used in real-world environments, covering user interactions, resulting outputs and impacts. A significant concern is the possible mismatch between the environments or populations for which AI systems were designed and those in which they are deployed. This misalignment can undermine accuracy and fairness, posing potential democratic risks . Key governance priorities during deployment include robust monitoring, accountability mechanisms, and competition policies to address unintended outcomes. Continuous oversight and maintenance are essential to evaluate the system's performance and societal impacts, feeding insights back into the design and development cycle to improve reliability and reduce risks in an iterative manner.

Staged release and access policies: Access governance involves phased or conditional releases of models to mitigate risks and improve system performance before wide-scale deployment (Kembery, Bucknall & Simpson, 2024). This approach allows developers to identify and resolve issues iteratively while ensuring accountability. Mechanisms like regulatory sandboxes (foreseen under the AI Act) and vetted researcher access enable controlled testing environments where safety and compliance can be closely monitored. For instance, API-based access with fewer users and gradual scaling up of access followed by rigorous monitoring is one practical option. Mandatory know-your-customer (KYC) screenings to verify the client's identity can also be envisioned before granting them access to AI systems with significant misuse potential (UK Government, 2023b, 39). Open-source deployments, though valuable for innovation, could also require curation and safeguards, such as licensing conditions or restrictions on fine-tuning, to ensure responsible use (see Seger et al., 2023a).

Monitoring and oversight: Continuous oversight and monitoring is essential for ensuring AI systems are deployed responsibly. These governance structures include mechanisms such as incident monitoring, where AI deployers have to monitor and share information to regulators about serious AI incidents and misuse as well as mitigation measures (Uuk et al., 2024). Based on this information, regulators can evaluate AI deployment, document accidents and intervene as required. This helps not only risk mitigation, but also in steering AI development towards the most societally beneficial directions (Whittlestone & Clark, 2021). Monitoring necessitates adequate procedures to receive, investigate, respond to and redress complaints about use of AI. Regulatory bodies, such as the AI Office and the AI Board in Europe or the international network of AI safety Institutes help to operationalize the oversight. Whistleblower protections can further enhance this by encouraging employees to report concerns about systems without fear of retaliation. Moreover, organisational practices such as board-level risk committees, chief risk officers, and internal audit teams can be mandated to report on and mitigate societal risks.

KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

Liability and legal remedies: Clear legal frameworks should assign accountability for AI-related harms and ensure access to effective remedies, enshrined in the European Charter of Fundamental Rights. Legislative mechanisms must establish who is liable — developers, users, or vendors — and empower individuals or groups to seek redress when negatively impacted. Transparency remains a challenge, as opaque AI systems and lack of disclosure about AI use make it difficult for affected parties to prove discrimination under the traditional burden of proof rules. Establishing the right to an explanation for AI decisions, which arguably exists under GDPR, and ensuring accessible complaint mechanisms are critical for upholding trust, protecting rights, and providing justice for those harmed by AI systems. The EU AI Act addresses some of these issues by requiring disclosure of AI systems that directly engage with individuals as well as notifications about being subject to decision-making by high-risk AI systems. However, gaps remain, which has led to proposals such as the AI Liability Directive to ease procedural burdens for claimants against AI harms.

Competition and market legislation: Competition legislation seeks to prevent monopolistic practices and ensure fair competition in the AI market. Given the AI market concentration across a number of layers, antimonopoly tools — industrial policy, structural separations, public options and cooperative governance — could be utilized to facilitate competition (Narechania & Sitaraman, 2024). In Europe, regulations like the Digital Markets Act (DMA) and Digital Services Act (DSA) aim to address anti-competitive behavior and protect consumers against large online platforms. The AI Act also partly builds on the internal market and product safety legislation under the new legislative framework to ensure that only high-risk AI products deemed safe for consumers gain access to the EU market. Recently, even US regulators have brought up antitrust cases and proposed breakup of dominant technology companies such as Google and Microsoft. By fostering competition and regulating market dominance, these measures can only advance democratic ideals such as lessening concentration of power.

3.5. Democratic pillars in relation to AI lifecycle

The six fundamental pillars of democracy (participation, freedom, equality, knowledge, transparency, rule of law) can be mapped to the relevant stages of the AI lifecycle to identify which governance mechanisms are particularly critical to supporting these pillars. For instance, transparency is arguably most related to development practices that ensure explainability and open auditing of models. The table below provides a high-level view of how each pillar aligns with specific governance mechanisms at different stages of the AI lifecycle.

Pillar of democracy	Most relevant AI lifecycle stage	Key governance mechanisms
Participation	*Design*: Digital public infrastructure; Participatory design requirements	Investing into digital public infrastructure to allow more participatory and open design of systems in Europe. Mandate participatory design principles or other stakeholder interaction to define the objectives of AI systems, especially in potentially high-risk cases within the public sector.

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

	Deployment: Monitoring and oversight; Liability and legal remedies	Ensure participatory oversight processes of AI systems, such as inclusive platforms for public feedback and whistleblower protections. Legal pathways for citizens to lodge complaints against AI harms and effective redress policies.
Freedom	*Design:* Pre-design and procurement	Incentivizing equitable and socially beneficial use cases of AI and prohibiting systems based on extensive surveillance, social scoring or other questionable uses of facial recognition or biometrics.
	Data: Privacy governance	Maintain privacy, data protection and security safeguards over data collection, storage and processing in training of AI systems to ensure confidentiality of personal data and avoid leaks of sensitive information.
Equality	*Data:* Data quality standards	Set requirements and standards for data provenance, including the origins, representatives and quality of datasets, depending on the social use case.
	Development: Bias mitigation	Require systematic evaluations on bias and discrimination through different fairness metrics, documenting the differences between social groups in accuracy and content of models.
Transparency	*Development:* Explainability and transparency policies; Model audits and evaluation	Mandate not only technically transparent and interpretable AI models, but also ones whose operations and decisions are explainable to citizens and users at large. Require external audits of AI models performance, safety and social impacts during development, especially those deemed high-risk.
Knowledge	*Deployment:* Monitoring and oversight; Competition and market legislation	Maintain public oversight of AI systems, along with open, citizen-facing reporting mechanisms and sharing information of incidents and ensure sufficient resources for regulators to investigate complaints. Moreover, regulate social media and large platforms to protect the information environment against misinformation.
Rule of Law	*Deployment:* Liability and legal remedies; Competition and market legislation	Set clear legal frameworks for accountability and liability for AI harms (e.g., between providers and deployers), especially in the public sector, ensure access to effective remedies for individuals whose rights are violated by AI systems. Utilise competition legislation against monopolization of the AI ecosystems and anti-competitive behaviour of large platforms to safeguard democratic processes.

Table 1 – Most relevant stages of the AI lifecycle for each pillar of democracy.

Funded by the European Union

KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

It should be noted that this mapping is a generalized overview to aid conceptual understanding and may not fully account for regional, sectoral, or cultural differences. Given that democratic impacts actualize in practice, the deployment is a crucial stage for all of the pillars. Yet, AI lifecycle stages appear to have an uneven impact across the pillars. For instance, development is especially relevant for equality whereas deployment is accentuated with rule of law. There are also considerable interdependencies between the pillars and AI lifecycle as governance failures at early stages (e.g., biased data) easily cascade into downstream impacts. It should also be noted that different stages of the AI lifecycle are not isolated but part of an interactive process where decisions at 'later stages' can also retroactively affect 'earlier stages' (e.g., redesign based on deployment feedback).

Participation cuts more evenly across the lifecycle stages, especially the design and deployment. There should be avenues for citizen participation in the pre-design stages, as well as mechanisms for liability and effective remedy towards individuals whose rights have been violated by AI systems. Governance approaches to freedom are more difficult to map, but at least design and data emerge as relevant domains. They involve, e.g., setting the right incentives for non-intrusive AI systems during pre-design and procurement stages as well as privacy standards against use of personal and sensitive data[2] in training AI systems.

Equality appears to be especially relevant in model development, given this is the stage where algorithmic biases and discrimination are evaluated and mitigated. That being said, equality is also reflected in the data, where such biases often stem from. Knowledge relates most strongly to deployment through oversight mechanisms and tackling concentration of power to very large platforms through competition policies. Increased public knowledge through open publication of risks and accessible reporting avenues are particularly important if power is perceived as non-domination. Similarly, transparency is a rather holistic quality across the AI lifecycle. While independent oversight institutions during deployment can greatly enhance transparency, it appears perhaps most relevant within development through explainability and external model evaluations.

The rule of law is perhaps most singularly focused on AI deployment even if it arguably also relates to stringent safety standards and robustness as part of development. Clear legal accountability and liability policies are especially important to ensure that effective remedies can be provided fairly to people harmed by AI. This also requires careful monitoring and oversight of AI deployment and its effects. Moreover, market and competition policies can lessen the concentration of power across the AI value chain, thereby, helping to maintain democratic checks and balances.

Figure 3 summarizes these findings into an overall framework. Democratic AI governance can be seen as consisting of four layers of AI lifecycle that build on top of each other, with the pillars of democracy cross-cutting them. Yet as noted, some pillars are more closely connected to certain stages of the AI lifecycle than others.

[2] Under Article 9 of GDPR, special category data (i.e. sensitive data) is personal data that reveals a person's racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetics, biometrics, health, sex life or sexual orientation.

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

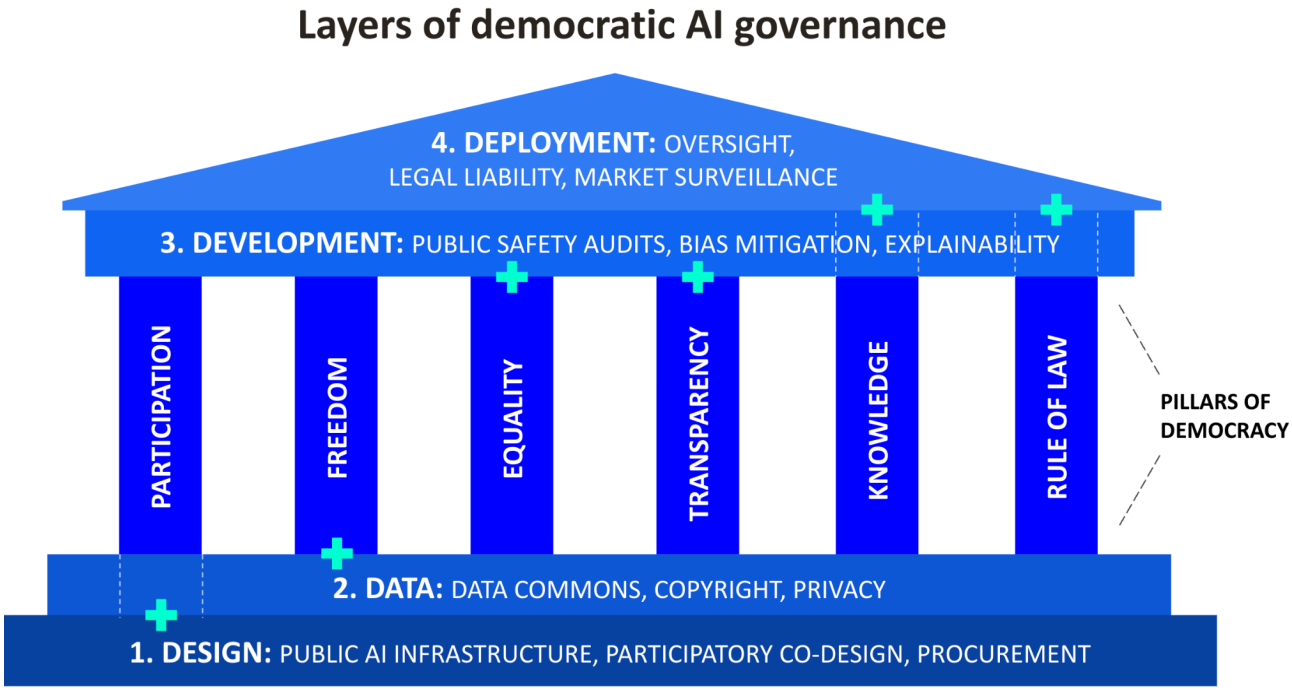


Figure 3 – How different layers of the AI lifecycle relate to pillars of democracy, with the strongest connections marked with a + sign.

This mapping of democratic pillars in relation to the AI lifecycle illustrates how deployment-oriented policies are perhaps the most important and opportune when safeguarding the democratic system. Indeed, deployment is the stage where the societal and democratic effects of AI systems actualize in practice, and values such as accountability and knowledge come to light. However, deployment is not an end all, and merely focusing on it risks disregarding earlier stages of the AI lifecycle, where important design and development decisions are already made. This includes the values and objectives steering the development of AI systems in the first place, such as equality. As emphasized by many (e.g., Selbst el al., 2019), AI systems should be understood as sociotechnical systems, where the technology interacts in tandem with the social context in which it is deployed. Democracy also involves a procedural perspective, which is challenging to realize if all of the digital infrastructure that facilitates the modern public sphere is owned by private platforms and large technology companies with profit incentives. Therefore, when assessing the democratic implications of AI, there is a need to holistically evaluate the entire lifecycle of AI systems. Democratising AI should be understood comprehensively to cover the use, development, profits and most importantly, governance of AI (Seger et al., 2023b).

Moreover, this indicates one should also consider how these governance mechanisms feed back to affect pillars of democracy. For example, only relying on stringent compute governance and safety standards could ultimately lessen the space from democratic engagement and deliberation. This is particularly true given the current dynamics of national safety and securitization around AI technologies, as these domains remain rather hostile to democratic input (Mügge, 2023). More general oversight measures, such as increased transparency and auditing of AI systems are more likely to support democratic participation. These dynamics are also somewhat visible in the current AI policy in the EU. For example, the AI Act is reliant on European standardization organisations to craft the technical standards by which companies can demonstrate compliance with the regulation of high-risk AI systems. However, these bodies tend to be dominated by the same private AI producers that the regulation targets, with limited civil society participation and democratic

KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

input (de Vries, Kanevskaia & de Jager, 2023). While empowering private standardization bodies and producers themselves with decision-making power about crucial aspects of AI regulation is a worrisome development, the European Commission has also shown willingness to engage with societal stakeholders at large. For instance, the drafting of the Code of Practice on General-Purpose AI models led by the EU AI Office involved over 1000 participants across industry, researchers and civil society (European Commission, 2025g).

3.6. Additional categorizations of AI governance

There are additional ways of categorizing and typologising AI governance approaches aside from the lifecycle view presented above. For example, one distinction highlighted in the first section is between **risk-based and rights-based approach** to AI governance. The risk-based approach is by its nature future-oriented and focuses on quantifying, assessing, and mitigating potential risks associated with AI systems. While there are different variations of risk-based regulation (for example, whether the risk assessment happens ex-ante versus ex-post), they typically involve assigning compliance responsibilities to AI developers or users to prevent foreseeable harms. This approach is enshrined in the EU AI Act, in which obligations are dependent on the different levels of risks involved with AI systems in proportional manner: minimal, limited, high and unacceptable risks. While systems with minimal risks are not subject to specific obligations under the AI Act, high-risk systems face more stringent requirements to be brought to markets. The focus on dynamic future risks and iterative risk assessments appears sensible given the rapid progress in AI and lack of clarity on the technological trajectory.

However, constructing AI harms as risks also comes with its own policy baggage, as noted by Kaminski (2022). Risk regulation often assumes that the use of technology is desirable and manageable through proper oversight, which can sideline more fundamental questions about whether AI should be deployed in certain contexts in the first place. Risk-based governance arguably prioritizes quantifiable technological solutions over systemic societal considerations by focusing narrowly on identifiable and measurable harms. Worryingly, these diffuse and abstract harms are precisely what characterises democratic risks posed by AI. Kaminski (2022) also notes that risk-based frameworks often lack robust mechanisms for individual redress and tend to obscure policy trade-offs as purely technical decisions. These deficiencies are particularly visible in the context of migration. By treating the fundamental rights of non-EU citizens as risks to be 'managed' through technical mitigations rather than protected through absolute rights, the AI Act arguably facilitates the continued use of invasive technologies in asylum and border processes.

On the other hand, rights-based governance is anchored in protecting fundamental rights, such as privacy or non-discrimination. The General Data Protection Regulation (GDPR) exemplifies this approach by prioritizing individual rights related to data protection, such as the right of access, rectification, erasure, data portability and to restrict processing. Rights-based governance frameworks are grounded in normative principles that establish universal standards of protection, regardless of the specific context or level of risk. As such, they might be better placed to address concrete harms of AI systems, for example by emphasizing the rights of marginalized groups that could be disproportionately affected by AI systems. Yet, a potential limitation of rights-based approaches is their rigidity, as they may impose uniform obligations that do not account for contextual nuances or variations in risk.

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

Of course, even if the risk and rights-based approaches are presented as distinct paradigms, they can overlap and complement each other. A hybrid governance framework might use risk assessment as a tool to operationalize the enforcement of fundamental rights. For instance, the EU AI Act incorporates some rights-based principles (e.g., non-discrimination and fairness) into its risk-based structure, demonstrating how these approaches can intersect with each other. Naturally, how the harms or risks of AI systems are defined and conceptualized in the first place is of significant importance. For example, Bommasani et al. (2021) identify risks of foundation models ranging from their technical principles (e.g., model architectures, training, data, security) and capabilities (language, vision, reasoning, human interaction) to their application areas (healthcare, education, law) and societal impacts (misuse, inequity, economic and environmental impact).

Other higher-level distinctions also exist, such as between **principles-based and rules-based approaches** to AI governance (Schuett et al., 2024). Principles-based approaches, such as those found in the OECD AI Principles (2019) or UNESCO's Recommendation on the Ethics of Artificial Intelligence (2021), emphasize flexibly implementing values like transparency, fairness, and accountability, but this tends to lack legal clarity. More enforceable are rules-based approaches, which detail precise obligations for developers, deployers, and users of AI systems, such as the AI Act, but which can become outdated in the face of technological progress. Another typology divides AI governance approaches into **sector-specific and cross-sectoral frameworks**. Sector-specific regulations, such as those governing AI in healthcare or autonomous vehicles, address the unique challenges and risks associated with particular domains, but threaten to create regulatory fragmentation. Cross-sectoral frameworks aim to provide overarching principles or rules that apply across multiple industries and promote harmonization but may lack the granularity required for specific applications.

Lastly, distinction can be drawn between **top-down versus bottom-up approaches**. Top-down approaches involve centralized regulation by governments or international bodies to ensure consistency but tend to be slow to develop and implement. Bottom-up approaches, on the other hand, emphasize decentralized and agile initiatives led by industry standards bodies, academic institutions, or civil society organisations (e.g., IEEE's voluntary guidelines and standards), but lack enforceability. Bottom-up approaches also connect to broader discourse on participatory AI governance (Wong et al., 2022), which emphasizes the involvement of diverse societal actors in shaping AI policies. Scholars like Buhmann & Fieseler (2023) emphasize the importance of "deep democratic" deliberation — structured, well-informed public debates that involve diverse stakeholders in critically evaluating the goals and functioning of AI systems. For instance, participatory governance efforts could expand beyond traditional user-centered design to prioritize community and societal perspectives, aligning with calls for human-centric AI governance that fosters emancipatory technology development (Sigfrids, 2023).

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

4. Building a culture of democratic AI governance

This section translates the democratic AI governance framework developed in the last section into concrete organisational practices. Each governance mechanism described below contributes to safeguarding one or more democratic pillars — participation, freedom, equality, knowledge, transparency and the rule of law — within the organisational context of AI development and deployment. The analysis builds on a legal perspective, focusing on key questions for companies and other organisations under the AI Act's risk-based approach. Nevertheless, it invites thinking in terms of crossovers and complementary angles for democratic affordances within the current regulatory and policy landscape.

As discussed, the AI Act (Regulation (EU) 2024/1689) establishes a comprehensive legal framework for trustworthy AI, built on risk-based obligations. However, compliance alone is insufficient to ensure that AI genuinely supports democratic values such as fairness, accountability, and inclusiveness. These values must be operationalised within organisations through internal governance systems that embed ethics and civic responsibility in the entire AI lifecycle. As such, this section translates the AI Act's regulatory obligations into practical guidance for AI developers and policymakers to promote democratic governance through AI.

4.1. Organisational foundations under the AI Act

Understanding Roles and Risk Levels under the AI Act

The AI Act aims to regulate the use of artificial intelligence within the EU Member States and set a common framework for the use and supply of AI systems and general-purpose AI models in the EU. It distinguishes several categories of AI systems — unacceptable, high-risk, limited-risk and minimal-risk — according to the level of risk they pose for individuals. It also recognises general-purpose models that may have either systemic or non-systemic risk. It further identifies four main actors: providers who design and develop systems; importers, who introduce them into the EU market; distributors, who make them available within the supply chain; and deployers, who operate them in practice.

Identifying AI Systems, their Risk Levels and Organisations' Role

The compliance process under the AI Act begins with verifying whether an organisation's activities, products, or services fall within the Act's scope. This requires confirming first whether a system is indeed an AI System, identifying and classifying it by risk level and determining the organisation's role — such as provider, distributor, importer, or deployer. Each role carries distinct legal responsibilities, with this section emphasizing the role of developers, who act as providers when they design and develop such AI systems.

Research and Development Exemption

It is important to first note that the AI Act excludes from its scope AI systems and models that are developed and put into service solely for scientific research and development (Article 2(6) AI Act), and it also excludes research, testing, and development activities carried out prior to an AI system being placed on the market or put into service (Article 2(8) AI Act). However, it does not cover real-world testing, nor does it remove compliance obligations once a system transitions toward deployment or market placement. In practice, this

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

means that experimental or exploratory work conducted in controlled research environments may proceed without triggering the Act's requirements, but the moment the system is intended for operational use, piloting in real conditions, or wider distribution, the relevant obligations apply in full. This exemption aims to preserve scientific freedom and support innovation (Recital 25 AI Act). In any case, all research and development activities must continue to respect applicable Union law and adhere to recognised ethical and professional scientific standards.

4.2. Important aspects of democratic AI governance

AI Literacy: Training and Awareness

From a democratic governance perspective, AI literacy is a foundational enabler of the democratic pillars of knowledge and participation. Equipping individuals within organisations with the ability to understand how AI systems function, assess their limitations, and critically evaluate automated outputs strengthens epistemic agency and supports meaningful participation in decision-making processes involving AI. Without adequate AI literacy[3], governance risks becoming purely technocratic, limiting the capacity of individuals to question, challenge or influence AI-driven outcomes.

Organisations should be committed to fostering a culture of accountability and preparedness by equipping its workforce with the knowledge and skills required to uphold the compliant use of AI technologies. To achieve this, organisations must promote AI literacy through training programs tailored to the role and expertise of each participant- whether team leaders, developers, or operators- and reflect the technical and operational context of the AI Systems used. Reviewing AI outputs require specialized instruction to identify risks, detect bias, and ensure both regulatory and internal compliance. AI literacy programmes should also reinforce democratic values (e.g., non-discrimination, fairness, transparency, accountability) and empower staff to identify democratic harms such as manipulation risks, opaque decision-making, or exclusionary data practices.

It is worth noting that recent policy discussions, including the proposal under the Digital Omnibus on AI regulation, suggest revisiting Article 4 of the AI Act by placing primary responsibilities for promoting AI literacy on EU-level institutions, such as the European Commission and the AI Office. This evolution reflects a growing recognition of AI literacy as a public governance function and a systemic prerequisite for democratic oversight of AI. However, even in this evolving regulatory context, organisational AI literacy remains a critical complement to institutional efforts. Without adequate internal understanding of AI systems, risks, and limitations, organisations would lack the capacity to operationalise democratic safeguards in practice, rendering high-level literacy initiatives ineffective at the point where AI systems are actually designed, deployed, and used.

Compliance with Copyright Law

Compliance with copyright and intellectual property law supports the democratic pillars of rule of law and equality, by ensuring that AI development respects legally protected interests, avoids the unlawful appropriation of creative works, and preserves fair conditions within the information and cultural ecosystem.

[3] While this deliverable focuses on AI literacy, the KT4D project has also utilised 'Critical Digital Literacy' as a more holistic concept. Please refer to the Module F of KT4D Social Risk Toolkit.

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

These democratic safeguards must be reflected in concrete organisational practices governing how training datasets are sourced, licensed and used throughout the AI development process.

In addition to the requirements of the AI Act and in line with existing EU law, an organisation's developers and policy makers (when acting as Providers) must ensure compliance with applicable copyright and intellectual property rights when developing or training their systems. This includes, in particular, the lawful sourcing and use of datasets, especially where such datasets may include content protected under copyright or related rights as well as obtaining the appropriate authorisation, licensing arrangements, or valid exceptions under applicable law.

Compliance with Data Protection Law

Compliance with data protection law is a cornerstone of democratic AI governance, directly supporting the democratic pillars of freedom, rule of law, and equality by limiting intrusive data practices, preventing arbitrary or disproportionate processing of personal data, and safeguarding individuals against power asymmetries inherent in large-scale data-driven systems. To give effect to these democratic principles, organisations must embed data protection requirements into the design, development and training of AI systems, rather than treating them as ex post compliance obligations.

Moreover, where an AI system processes personal data, organisations must also ensure full compliance with the GDPR and applicable data protection law, including national rules where relevant.[4] This includes clearly identifying the categories of data involved-such as names, contact details, or behavioural data-understanding who has access to the data, including internal teams and any external third parties, and assessing whether any personal data is transferred outside the European Economic Area, ensuring that appropriate safeguards are in place where required. This includes, where necessary, the performance of a Data Protection Impact Assessment ("DPIA") pursuant to Art. 35 of the GDPR.

If personal data is used to train the AI system, further precautions are necessary. Whether the data is collected directly by organisations from the individuals or received from third parties, organisations must ensure that:

- the origin and source of the data are traceable and properly documented;
- data subjects have been appropriately informed about the processing, in line with Articles 13 and 14 GDPR;
- a valid legal basis supports the processing, in accordance with Article 9 GDPR for special categories of data (such as health data) or Article 6 GDPR for non-sensitive (common) personal data.

Where obtaining consent is not feasible in practice, legitimate interest may be used as an alternative, provided that the processing pursues a genuine and specific interest, it is strictly necessary for that purpose, and a balancing test ("legitimate interest assessment" or "LIA") confirms that the data subjects' rights do not override that interest.[5] Additionally, if an organisation plans to reuse personal data it originally collected for

[4] As mentioned above, please refer to this link for the CNIL recommendations on GDPR compliance when organisations develop AI Systems – In French. Please also refer to the PDF file available for download here.
[5] For more details, please refer to EDPB's opinion on the use of personal data for the development and deployment of AI models (24th December 2024).

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

a different purpose, and data subjects were not previously informed, it must assess whether the new use is compatible with the original one. In all scenarios, organisations must comply with applicable transparency, documentation, and accountability obligations, and be able to demonstrate that any further use of data for AI training is lawful, proportionate, and technically safeguarded.

4.3. Practical guidelines for implementation of high-risk AI systems

An illustrative example of a high-risk AI system from a democratic perspective is represented by AI systems intended to influence the outcome of an election or referendum, or the voting behaviour of natural persons, as referred to in Annex III, point 8(b) of the AI Act and clarified in Recital 62. Such systems may be used in the context of political campaigns to analyse data relating to voter behaviour or preferences and to generate, select or distribute political messages or content to which voters are directly exposed. While these systems are not prohibited per se, their direct interaction with voters' decision-making processes places them in a particularly sensitive area of democratic governance.

This category of high-risk AI systems illustrates how democratic risks may arise even in the absence of manifestly unlawful or prohibited practices. Where AI systems are capable of shaping political information flows, influencing voter engagement, or affecting the formation of political preferences at scale, traditional compliance approaches focused solely on individual rights or technical performance prove insufficient. Instead, such systems require robust governance measures capable of operationalising democratic safeguards throughout the AI lifecycle, including at the organisational level where these systems are designed, developed, deployed and monitored.

Phase 1: Design and Development of the AI System

The foundation of compliance for high-risk AI systems begins during the design and development phase, where providers must integrate risk mitigation, data governance, human oversight, and cybersecurity considerations into the architecture of the AI system from the outset.

Risk Management System

From a democratic perspective, lifecycle-wide risk management systems operationalise the pillar of the rule of law, by ensuring that AI-related risks to fundamental rights are identified, documented, mitigated and traceable, rather than addressed in an ad hoc or discretionary manner. A continuous, lifecycle-wide risk management system (Article 9 AI Act) must be established to identify, analyse, and mitigate risks – including those affecting vulnerable groups and foreseeable misuse. This consists of fully documenting all processes and decisions taken during the system's development.

Data Governance and Quality Assurance

Robust data governance and bias mitigation measures are essential to uphold the democratic pillar of equality, as discriminatory outcomes often originate from unrepresentative, biased or historically skewed datasets. Data used to train, validate, and test AI systems must meet high standards of quality and governance (Article 10 AI Act). Input data, which may include personal data, sensitive information, or third-party records, must be managed in accordance with applicable confidentiality, security, and data protection obligations. Given this dependency, access to high-quality input data is essential to ensure performance and

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

prevent risks. Providers are required to implement documented procedures to ensure that all datasets are relevant, sufficiently representative, and free from unjustified bias (Recital 67 AI Act). Special attention is required to mitigate biases — whether present in historical datasets or emerging in real-world deployment — that could lead to discrimination prohibited by Union law (Recital 67 AI Act). Proactive inclusion of diverse and representative data is essential to support democratic equality. This is particularly important where feedback loops may cause outputs to affect future inputs, thereby amplifying existing disparities, especially those impacting vulnerable groups (Recital 67 AI Act).

Providers must maintain comprehensive documentation describing the system's general characteristics, capabilities, and limitations, as well as the algorithms and processes used for training, including how datasets are structured, tested and validated. System performance must be assessed both before deployment and on an ongoing basis, as well as properly documented for traceability and compliance verification throughout the system's lifecycle (Recital 71 AI Act). This includes implementing logging functionalities to enable automated recording of significant operational events (Article 12 AI Act).

To ensure ongoing performance, the design of the system must also support accuracy, robustness, and cybersecurity (Article 15 and Recital 76 AI Act). Technical solutions must be adopted that are appropriate to the relevant risks and circumstances across the entire lifecycle of high-risk AI systems, such as effective authentication, access controls and audit logging mechanisms to safeguard against unauthorised interference (Recital 74 AI Act). Furthermore, organisations shall implement robust monitoring procedures and response protocols designed to detect, assess and mitigate system-level vulnerabilities and exploitation attempts. All related documentation must remain clear, comprehensive, and up to date, in accordance with the applicable quality and regulatory requirements (Recital 66 AI Act).

Effective human oversight mechanisms support both the democratic pillars of freedom and rule of law, by preventing automation bias, enabling meaningful contestation of AI outputs, and preserving human agency in decision-making processes. In order to operationalise these democratic safeguards in practice, mechanisms to enable the deployer (Article 26(2) AI Act) to implement effective human oversight of the AI system must be in place (Article 14 AI Act). These mechanisms, including appropriate human-machine interface tools, should allow the deployer to monitor system performance, intervene when necessary, and override outputs, with a view to prevent or minimise risks to health, safety or fundamental rights that may emerge when the AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse. Human oversight measures must be commensurate with the risks, level of autonomy and context of use of the AI system and may include, inter alia, measures that enable interpretability of the system output and that avoid automation bias (Article 14(4)(b) AI Act).[6]

Phase 2: Pre-Market Readiness and Documentation

Before an AI system can be placed on the market or put into service, providers must demonstrate compliance through comprehensive documentation and ensure the system can be used safely by deployers. Technical documentation must be prepared in advance (Article 11 and Annex IV), detailing how the system satisfies the legal requirements required for phase 1, including risk management, data quality, human oversight, and other core areas. In addition, clear and accessible instructions for use must be provided to deployers (Article

[6] That is, the tendency of automatically relying or over-relying on the output produced by a high-risk AI system.

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

13 AI Act), explaining how the system should be operated, its limitations, potential risks, and oversight needs. These instructions also need to describe the data used and provide sufficient guidance for users to understand and control the AI system's behaviour. Providers are also required to establish a quality management system (QMS) (Article 17 AI Act). This includes policies and procedures for compliance management, system design and validation, data governance, and risk controls. The QMS must also cover post-market monitoring, incident reporting, stakeholder communication, and internal accountability.

Phase 3: Conformity Requirements for Market Placement

Before placing a high-risk AI system on the EU market or putting it into service, providers must complete the AI Act's conformity assessment procedure (Article 43 AI Act), issue the EU Declaration of Conformity (Article 47 AI Act), and apply the CE marking (Article 48 AI Act). In addition, the system must be registered in the EU database for high-risk AI systems (Articles 49 and 71 AI Act), including key information such as its intended purpose, conformity assessment route, and declaration reference (Annex VIII AI Act).

Phase 4: Post-Market Monitoring and Incident Management

Compliance does not end once a high-risk AI system is placed on the market. Providers are required to implement a post-market monitoring system to ensure the system continues to perform safely (Article 72 AI Act). This monitoring system must be proportionate to the nature and risks of the AI system and should collect and analyse relevant data throughout the product lifecycle. It should incorporate feedback from deployers, real-world performance insights, interoperability issues, and other operational data. Providers must also establish internal processes to identify, assess, report and investigate serious incidents involving high-risk AI systems (Article 73 AI Act). Moreover, corrective actions are required (Article 82 AI Act) even where the system technically complies with the AI Act after having performed an assessment, if new risks emerge that threaten fundamental rights or safety.[7] Corrective measures may include updating, disabling, withdrawing the system, or limiting its functionality to address the issue effectively.

4.4. Guidance for limited-risk AI systems

An example of a limited-risk AI system with significant democratic implications is the use of AI-generated or AI-manipulated content on matters of public or civic relevance, where individuals are directly exposed to the system's outputs, such as synthetic audio, video or images circulated through online platforms. This could include, for instance, deepfake content depicting political candidates, public officials or political events that circulates on social media in the period immediately preceding elections or referenda, without clear disclosure that the content has been generated or manipulated by AI. While such systems do not fall within the category of high-risk AI systems under the AI Act, their impact on democratic processes is explicitly addressed through the transparency obligations set out in Article 50. These obligations are intended to

[7] According to Article 79(1), a "product presenting a risk" is defined in Article 3(19) of Regulation (EU) 2019/1020 and includes any system that may endanger health, safety, or fundamental rights. As per Article 3(19) of Regulation (EU) 2019/1020, a 'product presenting a risk' means a product having the potential to affect adversely health and safety of persons in general, health and safety in the workplace, protection of consumers, the environment, public security and other public interests, protected by the applicable Union harmonisation legislation, to a degree which goes beyond that considered reasonable and acceptable in relation to its intended purpose or under the normal or reasonably foreseeable conditions of use of the product concerned, including the duration of use and, where applicable, its putting into service, installation and maintenance requirements.

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

preserve individuals' ability to understand the nature and origin of the content they are exposed to, thereby safeguarding transparency, epistemic agency and informed participation in public debate. In the absence of effective disclosure and marking measures, even limited-risk AI systems may contribute to the distortion of the information environment, undermine trust in democratic institutions and exacerbate asymmetries of power in the public sphere.

In this context, transparency obligations regarding human–AI interaction and AI-generated content directly reinforce the democratic pillars of transparency and knowledge, by enabling individuals to understand when algorithmic systems shape information, communication, or decisions affecting them.

Human-AI Interaction Disclosure

An AI system intended to interact directly with humans must be designed in a way that ensures natural persons are informed that they are directly interacting with an AI system, unless this is clearly obvious to a reasonably well-informed, observant, and circumspect user, considering the specific circumstances and context of use (Article 50(1)).

Marking AI-Generated Content

AI-generated or manipulated content (synthetic audio, image, video, text) must be marked in a machine-readable format as artificial and embed technical solutions that enable detection that the output has been generated or manipulated by an AI system and not a human (Article 50(2)). When implementing that obligation, the characteristics of natural persons belonging to vulnerable groups due to their age or disability should be taken into account to the extent the AI system is intended to interact with those groups as well (Article 50(1) and Recital 132). Exemptions for this marking obligation exist, whereby AI systems performing primarily an assistive function for standard editing or AI systems not substantially altering the input data provided by the deployer or the semantics thereof are not covered (Article 50(2) and Recital 133). Below is a sample notice or disclosure, to be tailored according to the AI system and its specific functionalities:

> ***Important Notice: Interaction with an AI System***
>
> *You are interacting with a system that uses artificial intelligence (AI) technologies.*
>
> *This system operates autonomously, without real-time human oversight, and performs its functions using AI-based logic and processing. While designed to support various activities efficiently, it may produce outputs that are incomplete, imprecise, or not fully aligned with specific circumstances.*
>
> *The system does not possess independent decision-making authority or the ability to understand human context. Its functioning is limited to the scope of its design and training, and it may not account for all relevant variables or nuances.*
>
> *You are therefore advised to use the system as a support tool only. It is not intended to replace professional judgment or consultation with appropriate colleagues or competent business functions, particularly in relation to important or sensitive decisions.*
>
> *This requirement reflects the principle of transparency and aims to prevent unintentional deception, particularly in interfaces that may closely resemble human interaction or involve decision support.*

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

As the enforcement of the AI Act and EU's digital regulation remains politically contentious, it is yet to be determined how these democratic affordances materialise. Ultimately, the realisation of democratic AI governance depends on a cultural shift that moves beyond the binary of innovation versus regulation. This would mean AI developers approaching values like the rule of law and public participation not as administrative burdens, but as essential mechanisms for ensuring public trust in the technology. However, embedding ethics within individual organisational workflows is only one part of the equation. To ensure these democratic values are protected at scale, organisational efforts must be anchored in a systemic, long-term policy vision. The following section translates these conceptual and organisational requirements into a concrete EU policy roadmap and set of recommendations designed to safeguard the democratic sphere through 2035.

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

5. EU policy roadmap and recommendations

This section presents the policy roadmap and recommendations based on the conceptual framework presented in the earlier sections. These policies support the realisation of the framework for democratic AI governance, guiding European policymakers in governing AI systems to reinforce democracy from a holistic, infrastructural perspective. The section consists of:

- Policy categories
- Policy roadmap
- Policy recommendations

The categories are clusters of AI policies, marking areas or domains of interventions that the policy roadmap and recommendations focus on. The roadmap is organised into sequences of short, mid and long-term policy actions, which are foreseen to support the goal of AI strengthening democracy by 2035. The recommendations specify which kind of decisions and interventions should be taken in the light of the ongoing and anticipated policy processes, to ensure that the roadmap follows through. While stakeholder engagement has formed important inputs into the policy recommendations and the roadmap, their final format is the result of overall analysis of data, including review of relevant literature.

Methodologically, this section draws from the three workshops organised as part of the KT4D's Use Case 1 in Brussels, featuring policymakers, civil servants, researchers, think tanks and NGO representatives. The last use case meeting organised in November 2025 in Brussels (see D1.2V3) was specifically focused on co-creating the roadmap. The recommendations also incorporate findings from a Delphi study on the future of AI governance organised between April 2024 and June 2024 as part of KT4D. The two-round Policy Delphi consisted of 29 European AI experts, covering questions about the likelihood and desirability of different risks, trends and priorities in AI policy. There were six main categories of questions in total, covering 1) expectations regarding risks of AI, 2) implementation of the European AI regulation, 3) citizen participation, 4) global cooperation, 5) future-proof regulation, and 6) Industrial policy for AI in Europe. The Delphi study identified a desirability-probability gap, meaning that desirable AI policy directions, such as greater citizen participation, were perceived as less feasible. Moreover, the result emphasized the role of practical implementation and enforcement for future-proof AI regulation (see preprint publication here).

5.1. Policy categories

The policy categories identify the key levers and instruments of AI policymaking for implementing the roadmap and its recommendations. By clarifying where interventions are needed, they provide a way to operationalise the governance framework, which brings together the pillars of democracy and the AI lifecycle. As such, the categories set out the essential policy areas where sustained attention, capacity, and investment are needed over time. The roadmap, introduced after the policy categories, by contrast, introduces dynamism. It sequences how these categories are activated, strengthened, and connected across different time horizons.

The five policy categories are as follows:

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

1. **Regulatory enforcement:** This category draws attention to the implementation of the EU's recent digital and AI regulatory framework, following the first Von der Leyen commission's regulatory legacy and subsequent deregulatory developments. It centres on the question of how existing rules are enforced in practice, and with what capacities and resources. Attention is placed on both EU and national level, including public sector capabilities, effective enforcement of the AI Act, Digital Services Act and Digital Markets Act, mandatory transparency and disclosure requirements, robust fundamental rights impact assessments, and the integration of democratic considerations into enforcement practices.
    - Covers factors such as: Enforcement capacity and institutional strength, regulatory coherence and executability, anticipatory and global governance
2. **Public AI infrastructure**: This category addresses the need to build and govern shared European AI infrastructures that reduce structural dependencies on foreign technology providers. It includes initiatives such as Eurostack, AI factories and other European-level infrastructure projects, alongside investments into sovereign cloud and data centre capacity. Yet, it is also important that such digital public infrastructure is governed democratically, for instance through data or AI commons.
    - Covers factors such as: Sovereign infrastructure, public value and commons-based governance, sustainability and energy governance
3. **Investments and innovation**: Here the focus is on leveraging European AI investment and innovation policy towards strategic and democratic objectives. This includes increased AI investments under new multiannual financial framework, the development of the digital single market and capital markets, support for SME AI adoption, initiatives such as EU-INC, and the use of public procurement, such as buy European requirements, to advance open-source solutions and interoperability.
    - Covers factors such as: Capital mobilisation and financial coordination, single market, public procurement
4. **AI literacy and participation:** This category focuses on democratic participation that should underpin AI governance. It covers measures such as improved AI literacy through education and training[8], protection of electoral processes from misinformation, the meaningful involvement of civil society and digital rights organisations in AI development and policymaking as well as establishment of a citizens' assembly on AI governance.
    - Covers factors such as: Foundational AI literacy and skills, information integrity, Institutionalised participation and co-design
5. **Research and standards**: This category builds on all of the categories above, emphasising the role of research, experimentation, and standard-setting in shaping future-proof and democratic AI development. It encompasses sustained investment and commitment to experimental research and innovation on democratic AI under Framework Programme 10, creation of shared, high-ambition

[8] Enhancing citizen's critical digital literacy through education is one of the goals of KT4D and its KER4. Please refer to project tools such as: Serious game, Deepfakes interactive explainer, and Algorithms interactive explainer.

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

European research and infrastructure capacity for AI similar to CERN, focus on AI safety and security research, and active shaping of international standard-setting processes.

- Covers factors such as: Mission-driven research, interdisciplinary AI research, Standards, certification and auditability

5.2. Roadmap

The policy roadmap presented in this section guides the implementation of the framework on different timeframes, thereby bringing dynamic dimension to the policy categories. It builds on the results of the last use case meeting in November 2025 to translate the conceptual framework for democratic AI governance into a sequence of actionable policy directions. Rather than starting from existing policy instruments alone, the roadmap was developed through a structured backcasting exercise, beginning from a shared long-term vision in which AI has not undermined democratic systems but contributed to their strengthening by 2035. Participants were invited to reflect on the institutional, infrastructural, and normative conditions that would need to be in place for such a future to materialise, and to work backwards to identify critical milestones, dependencies, and policy interventions across short-, medium-, and long-term horizons. The resulting roadmap should therefore not be read as a prescriptive implementation plan, but as a collectively informed articulation of plausible pathways. It is structured around the policy categories described above.

Short-term 2026–2028: Enforcing regulation & building public AI infrastructure

- This phase focuses on defending and enforcing existing regulation while initiating development of public AI infrastructure. It recognises that democratic AI governance depends not only on rules and standards, but on digital sovereignty across the technical stack itself, alongside sufficient public-sector capacities to implement, oversee, and enforce regulation in practice.
  - Strategic objectives to be achieved by 2028: Operational enforcement network on AI regulation, online information ecosystem resilient to misinformation, EU standards adopted in global governance, investments and procurement align private and public AI

Mid-term 2029–2032: Democratic adoption

- As European public AI infrastructure matures, this phase emphasises democratic AI adoption. Democratic safeguards move from isolated requirements to default practices embedded across AI lifecycle through mature institutions, aligned incentives, and broad public understanding. Democratic practices at this stage ensure that the public infrastructure and regulatory frameworks are not co-opted by purely private interests.
  - Strategic objectives to be achieved by 2032: Widespread public understanding of AI, institutionalised participation and co-design, sovereign AI infrastructure fully operational, anticipatory governance mechanisms on AI

Long-term 2033–2035: Exercising AI sovereignty

- This phase marks the point at which public AI infrastructure and democratic practices converge, enabling long-term sovereignty and sustained public agency in shaping the development and use of

KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

AI. It centres on establishing AI sovereignty with the built infrastructure and democratic practices, understood as the ability of public institutions and citizens to meaningfully govern critical AI infrastructures, data, and capabilities rather than remaining dependent on external or purely commercial actors. It is only as public AI infrastructure and democratic participation meet that democratic AI governance becomes possible.

- Strategic objectives to be achieved by 2035: Ecosystem of leading AI companies based in Europe, democratic safeguards codified in technical standards, broad public oversight of AI development

A high-level overview of the roadmap is provided in Figure 4, outlining how the vision of stronger democracies through AI could be actualised across three successive timeframes. It illustrates the cumulative logic of change from the short-term to the long-term, linking what needs to change in terms of institutional capacity, infrastructure, and societal capabilities. These AI policy actions are further explicated in the roadmap table below as well as in the following recommendations section.

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

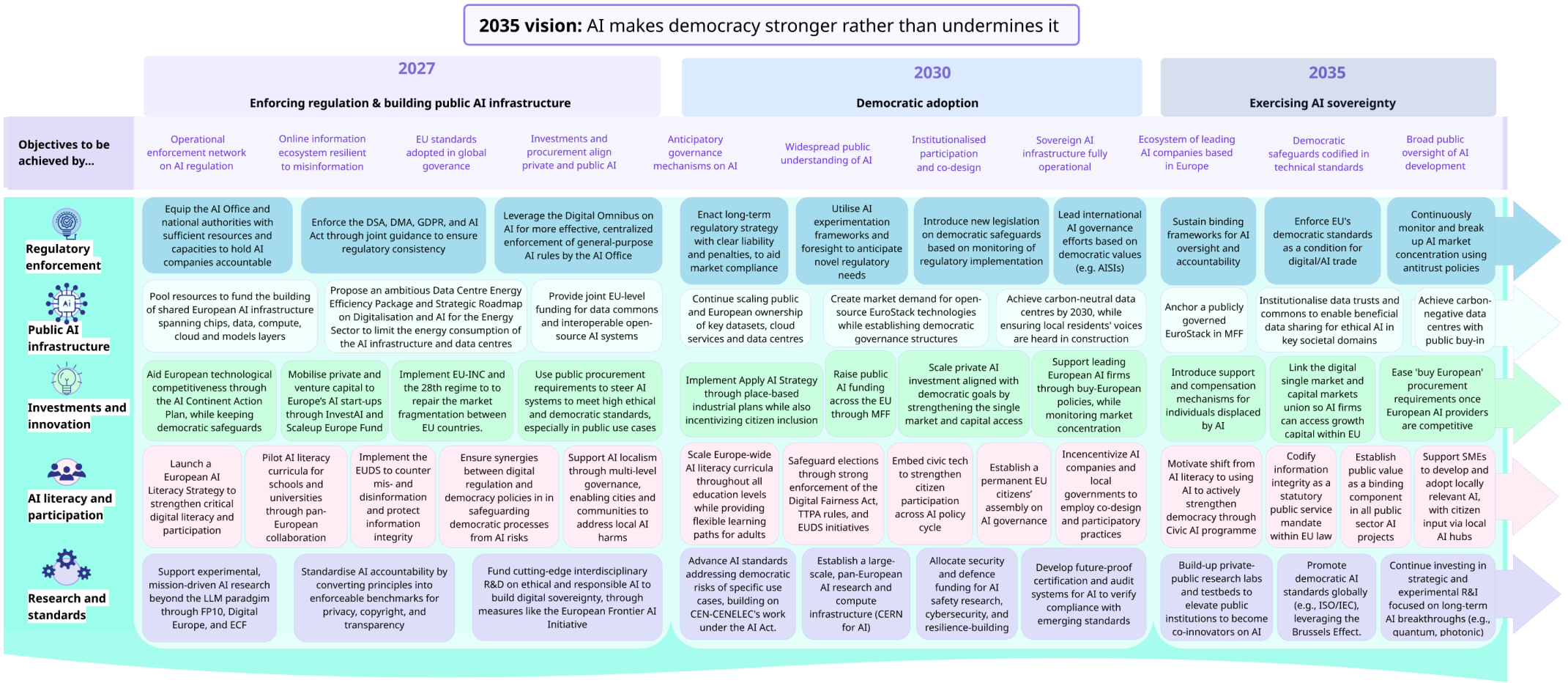


Figure 4 – Visualisation of the policy roadmap, detailing policy actions to be taken between 2026 and 2035 to advance democratic AI governance in Europe.

Funded by the European Union

KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

2035 vision: AI makes democracies stronger rather than undermines it

2026-28	2029-32	2033-35
Regulatory enforcement		
→ **Build regulatory enforcement capacity by providing sufficient resources**, training and tools for the Commission's AI Office and national authorities to hold AI companies accountable and protect democratic values, as full enforcement powers of the AI Act kick-in 2026 and 2027. → **Enforce the DSA, DMA, GDPR and AI Act (including Code of Practice) in coordination through joint guidance and supervision** to streamline overlapping obligations and ensure these rules apply consistently across platforms and AI systems. → **Leverage the Digital Omnibus on AI for more effective, centralized enforcement of general-purpose AI rules** by the AI Office (including algorithms by VLOPs/SEs) and expanded pre-market conformity testing of high-risk systems.	→ Establish a sustainable AI regulatory strategy to **strengthen market predictability and compliance via transparent liability rules** and penalties. This should integrate system interoperability and reliability standards, while ensuring coherent restrictions on non-compliant foreign AI systems. → Utilise **experimentation frameworks, such as regulatory sandboxes** under the AI Act, while embedding mandatory horizon scanning and technology assessment directly into the legislative process to anticipate emerging AI risks and regulatory needs. → **Monitor implementation of regulation, evaluate the need for amendments to the AI Act** and introduce new legislation on democratic safeguards as necessary, based on enforcement learnings and fundamental rights impact assessments, while also considering different obligations based on organisational size and capacity as market structures evolve. → **Lead global AI governance efforts and formalize international cooperation** with EU allies centred on democratic values, for example through standardisation and network of AI Safety Institutes (AISIs).	→ Sustain **legally binding frameworks that institutionalize public oversight of AI and accountability**, coupled with continuous civic and public administration capacity building. → **Re-examine and consolidate the EU digital regulatory architecture** to ensure long-term strategic coherence across different policy instruments, e.g. in terms of risk-assessment obligations and data-sharing mandates. → **Formalise a distinct democratic approach to AI governance,** using the Brussels effect to enforce EU standards as a condition for global trade, countering the "Washington effect". → **Monitor and break up market concentration** continuously using new antitrust and foresight mechanisms to mitigate emerging oligopolistic tendencies across the AI value chain.
Public AI infrastructure		
→ Fund the building of **shared European AI infrastructure through pooled resources, following the Eurostack model**, including chips, data, compute, cloud, and model layers to reinforce EU's digital sovereignty.	→ Continue investing in, operationalizing and **scaling public ownership of key datasets, cloud services and data centres across** Member States and network of European AI hubs to enhance democratic resilience, building on the basis of Common European data spaces.	→ Move from project-based investments to a **permanent, publicly governed EuroStack, with long-term budget commitments** anchored in the Multiannual Financial Framework (MFF).

Funded by the European Union

KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

→ Ensure **joint EU-level funding for development of data commons, interoperable open-source AI systems** and scaffolding as part of European Open Digital Ecosystem Strategy to reduce dependency on private, proprietary systems. → Propose an ambitious Data Centre Energy Efficiency Package and Strategic Roadmap on Digitalisation and AI for the Energy Sector to **limit the energy consumption of the AI infrastructure and data centres** to a sustainable, net zero trajectory.	→ Create guaranteed **market demand for open-source EuroStack technologies while also establishing democratic governance structures** for such public infrastructure. This is in order to steer the technology towards societal use cases and to ensure fair data labelling and labour practices. → Reduce the **environmental footprint of AI infrastructure and achieve carbon-neutral data centres by 2030,** in line with the data centre energy efficiency Package and associated EU frameworks, while ensuring voice of local residents in decision-making about data centre construction and grid capacity.	→ Institutionalize **democratic data governance models (such as data trusts or commons) for beneficial data sharing** to incentivize development of ethical, safe, and privacy-preserving AI systems in societal domains like health and transportation. → Achieve **carbon-negative data centres and cloud,** positioning Europe as a global leader in sustainable AI infrastructure with public buy-in.
Investments and innovation		
→ Aid European technological competitiveness while enforcing democratic safeguards by **implementing the AI Continent Action Plan initiatives** like the Apply AI Strategy, Cloud and AI Development Act and Data Union Strategy. → Mobilise **private equity and venture capital** for investments into the ecosystem of European AI start-ups, scale-ups and SMEs through initiatives like the InvestAI and Scaleup Europe Fund. → Advance the digital single market and unified capital markets in the EU through structures such as the **EU-INC and 28th regime** to repair the fragmentation between European countries. → Use **public procurement conditions to steer AI systems** to meet high ethical and democratic standards,	→ Implement the Apply AI Strategy at full scale through **place-based industrial plans and innovation ecosystems around AI**, building on the Experience Centres for AI / European Digital Innovation Hubs (EDIHs), while also introducing community-based conditionalities for transparency and citizen inclusion in their development. → Raise **public AI investment across Europe** through the next Multiannual Financial Framework (MFF) and fully operationalise the 28th regime as a unified European regulatory and operating environment. → Scale **private AI investment in line with strategic** and democratic goals by strengthening the digital single market, improving access to capital, and updating enabling frameworks such as bankruptcy law and skilled migration rules. → Continue to support the **growth of leading European AI companies through a mixture of "buy European" policies** and experimental research and innovation pathways, while also monitoring undemocratic market	→ Establish lasting **support, compensation and reskilling mechanisms** for individuals and **professions displaced by AI** to bridge innovation policy with just transition. → Complete the link between the **digital single market and capital markets union**, ensuring that European AI companies can access growth capital without relocating or restructuring outside the EU. → Gradually ease "Buy European" procurement requirements once **European AI providers are sufficiently established** and competitive against foreign hyperscalers, shifting focus to global exports.

Funded by the European Union

KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

especially in public use cases.	concentration.	
AI literacy and participation		
→ Launch a **European AI literacy strategy** under the European Democracy Shield (EUDS) to strengthen critical digital literacy, skills, and participation across population groups. → Develop and **pilot AI literacy curricula for schools** and universities through pan-European collaboration. → Implement the actions within the **European Democracy Shield to counter misinformation and disinformation** and safeguard the integrity of the information space. → Ensure synergies between the **digital regulation (AI Act, DSA) and democracy policies,** including the EUDS, European Media Freedom Act, Digital Fairness Act, and Transparency and Targeting of Political Advertising (TTPA) regulation in safeguarding free and fair elections from risks of AI. → Support **AI localism through a distributed, multi-level governance approach,** so that cities and local communities can play an active role in addressing governance gaps left from the EU-level.	→ Scale and **update a Europe-wide, mandatory AI literacy curricula** from primary education through secondary and higher education while also providing flexible, blended learning opportunities for adults by supporting local critical digital literacy initiatives by NGOs. → Safeguard democratic processes, including elections, through robust **implementation of the Digital Fairness Act and enforcement of TTPA rules**, as well as increased coordination capabilities of the European Centre for Democratic Resilience. → Develop and **embed civic tech that improves the scale and quality of citizen participation** in democratic innovations used by EU institutions, while adhering to Better Regulation guidelines throughout AI policy cycle to ensure that initiatives are grounded democratic stakeholder engagement. → Establish a permanent **European citizens' assembly on AI governance** consisting of a representative sample of randomly selected EU citizens to guide decisions about acceptability of AI development and use. → Incencentivize **AI companies and local governments to employ co-design and participatory practices** to systematically involve local communities, citizens and digital rights organisations in the development of AI, in order to build public trust in such systems, e.g., through taxation, procurement or local AI sandbox policies.	→ Institutionalise a Civic AI programme that **shifts from AI literacy towards designing and using AI to actively strengthen democratic practices,** including facilitation of deliberation, educational tutoring and government transparency. → Codify **information integrity as a statutory public service mandate** within EU law**,** requiring broadcasters and platforms to maintain AI-assisted verification systems and decentralized fact-checking protocols as a condition of operation. → Make **'public value' a standard, binding component in all public sector AI projects**, to prioritise the deployment of AI systems that demonstrably enhance wellbeing, societal resilience and public service quality, rather than merely optimising for efficiency or engagement metrics. → Provide continuous support for **small and medium-sized enterprises (SMEs) to develop and adopt AI tools** that are relevant, and appropriate for the communities they serve, whilst also ensuring input of the affected citizens through local AI hubs.
Research and standards		

Funded by the European Union

KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

→ Support experimental and **novel research on emerging technologies that extends beyond the current LLM paradigm** through DARPA-like mission-driven innovation policy under the FP10, Digital Europe programme and European Competitiveness Fund. → Fund cutting-edge **interdisciplinary research and development on democratic and responsible AI** to build digital sovereignty, through cross-national measures like the European Frontier AI Initiative. → Standardize **AI accountability by converting broad regulatory principles into accessible, enforceable benchmarks** for privacy, copyright and transparency, fit for evaluating democratic implications of AI, building on the AI Act and GPAI CoP.	→ Allocate a portion of the increasing **security and defence funding for AI safety research**, cybersecurity, and resilience-building to counter AI's more systemic risks to economy, culture and democracy, in collaboration with networks like ELSA. → Establish and **govern a large-scale, pan-European AI research and compute infrastructure (CERN for AI),** pooling expertise, resources and funding across Member States to secure long-term scientific and technological sovereignty in AI, potentially coupling it with efforts to scale-up RAISE initiative under the next long-term EU budget for 2028-2034. → Further develop **future-proof certification and audit systems for AI,** establishing independent mechanisms to verify compliance with emerging standards, for instance through AI Office's enforcement of GPAI systems. → Advance the creation of **AI standards and guidance tailored to contextual democratic risks of specific use cases,** building on the work of CEN-CENELEC on harmonised standards under the AI Act.	→ Continue Investing in **strategic and experimental R&I focused on long-term AI breakthroughs,** covering quantum computing, photonic technologies, and other emerging approaches that move beyond silicon-based architectures. → Build-up **private-public research labs and testbeds** to elevate public institutions from passive downstream deployers to co-innovators on disruptive AI. This capacity building develops the technical skills required to audit emerging systems. → Promote **democratic AI standards globally beyond the EU (e.g., ISO/IEC)**, working with partners to ensure European norms are recognised and adopted globally, leveraging the Brussels Effect.

Table 2 – The overall policy roadmap divided into actions in short, medium long-term across five policy categories.

Funded by the European Union

KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

5.3. Recommendations

These recommendations showcase a broader, infrastructural lens to democratic AI governance. They couple investments into digital public infrastructure with robust enforcement of existing regulation to tackle concentration of power in AI. This is designed to more holistically cover the whole AI lifecycle, since focusing merely on deployment of AI systems disregards the earlier, democratically significant decisions. Democratic governance and trust is hard to realize afterwards if our digital infrastructure comes to be owned by private, profit-seeking platforms. The recommendations seek to make democratic values a tangible competitive advantage for the European AI sector, distinguishing it from unreliable, authoritarian approaches.

The recommendations deepen the policy roadmap's 2026-2028 actions further through additional governance interventions. As such, they seek to cover what should be done in the next 3 years from a policy perspective. The recommendations are intended for European policymakers working on AI policy and associated fields, but also provide some guidance for national and local stakeholders.

Regulatory enforcement:

1. **Build regulatory enforcement capacity by providing sufficient resources, training and tools** for the Commission's AI Office and national authorities to hold AI companies accountable and protect democratic values, as full enforcement powers of the AI Act kick-in 2026 and 2027.
   - 1.1. Invest in institutional capacity building, continue the AI Office's open consultations and encourage similar deliberative practices with other EU regulators.
   - 1.2. Secure the AI Office's in-house expertise for third-party testing and evaluations to enforce GPAI Code of Practice and to avoid regulatory capture by industry.
   - 1.3. Ensure sufficient contestation and redress mechanisms for individuals harmed by AI systems within existing legislation in the absence of the AI Liability Directive.

2. **Enforce the DSA, DMA, GDPR and AI Act (including Code of Practice) in coordination through joint guidance and supervision** to streamline overlapping obligations and ensure these rules apply consistently across platforms and AI systems.
   - 2.1. For instance, amend the systemic risk classification of GPAI models in AI Act to focus more their scale and reach, similar to the definition used in DSA for VLOPs/SEs (Art 34) to better cover risks to democratic participation
   - 2.2. Moreover, ensure that the fundamental rights impact assessments (FRIA) under AI Act are aligned and mutually recognized with data protection impact assessments (DPIA) of GDPR.
   - 2.3. Enforce the DMA and other antitrust rules against gatekeeper companies to prevent abuses of dominant market position.

3. **Leverage the Digital Omnibus on AI for more effective, centralized enforcement of general-purpose AI rules** by the AI Office (including algorithms by VLOPs/SEs) and expanded pre-market conformity testing of high-risk systems.
   - 3.1. Challenge and limit the effects of the proposed Digital Omnibus changes to automated decision-making rules under GDPR and the removal of EU database registration requirements for certain high-risk systems under Art 6(3) in the AI Act.

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

3.2. Improve instruments such as ad labelling and watermarking of AI-generated content based on the Code of Practice on Transparency of AI-Generated Content to safeguard the information environment.

3.3. Explore extending the AI Act's fundamental rights impact assessment requirements (Art 27) on high-risk AI systems beyond public bodies to cover democratic risks more holistically.

Public AI infrastructure:

4. **Fund the building of shared European AI infrastructure through pooled resources,** following the Eurostack model, including chips, data, compute, cloud, and model layers to reinforce EU's digital sovereignty.
   4.1. Advance the European Chips Act 2.0 to improve the EU's semiconductor ecosystem for advanced AI chips by coordinating actions with the Critical Raw Materials Act.
   4.2. Leverage the coordinated investments into AI factories, AI Antennas and European Data Spaces to include provisions for public use, thereby democratising access to data and computing power.

5. **Ensure joint EU-level funding for development of data commons, interoperable open-source AI** systems and scaffolding as part of European Open Digital Ecosystem Strategy to reduce dependency on private, proprietary systems.
   5.1. Finance and utilise European Digital Infrastructure Consortium (especially Digital Commons EDIC) in implementing multi-country projects for open, interoperable and scalable public AI infrastructure.
   5.2. Engage communities like the European Network for Technological Resilience and Sovereignty (ETRS) to create a detailed roadmap for digital sovereignty across industry, research, think tank and NGO stakeholders.

6. **Propose an ambitious Data Centre Energy Efficiency Package** and Strategic Roadmap on Digitalisation and AI for the Energy Sector to limit the energy consumption of the AI infrastructure and data centres to a sustainable, net zero trajectory.
   6.1. Build on and ensure coherence of energy efficiency requirements between the AI Act, recast Energy Efficiency Directive (EED), Taxonomy Regulation, European Code of Conduct for Energy Efficiency in Data Centres, common Union rating scheme for data centres, ecodesign requirements for servers and data storage products and EU Green Public Procurement (GPP) Criteria for Data Centres, Server Rooms and Cloud Services, etc.

Investments and innovation:

7. **Aid European technological competitiveness while enforcing democratic safeguards** by implementing the AI Continent Action Plan initiatives like the Apply AI Strategy, Cloud and AI Development Act and Data Union Strategy.
   7.1. Boost private, European investment in sustainable cloud and data architecture by enacting the Cloud and AI Development Act and Data Union Strategy.
   7.2. Prioritize Europe's competitive strengths, namely ethics, trustworthiness and privacy in AI

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

adoption among SMEs and small mid-caps as part of the Apply AI Strategy.

8. **Mobilise private equity and venture capital for investments into the ecosystem of European AI start-ups, scale-ups and SMEs** through initiatives like the InvestAI and Scaleup Europe Fund.
   8.1. Reserve multi-billion funds to invest in the most promising European companies in strategic areas of AI development and deployment in accordance with European values.
   8.2. Coordinate funding streams between European Innovation Council, European Investment Bank, Member State innovation funders, sovereign wealth funds and philanthropic VCs.
   8.3. Balance the hardened screening of foreign investment on dual-use equipment and critical technologies, whilst also ensuring foreign investment, innovation and positive spillover is not jeopardised.

9. **Advance the digital single market and unified capital markets in the EU through** structures such as the EU-INC and 28th regime to repair the fragmentation between European countries.
   9.1. Establish a standardized pan-European corporate structure with harmonised corporate governance, capital maintenance rules and online registry to aid startup/SME scaling and ease regulatory compliance across Member States under the EU–INC proposal.
   9.2. Facilitate early-stage funding and simplify cross-border operations by standardising investment processes and employee stock options.

10. **Use public procurement conditions to steer AI systems to meet high ethical and democratic standards**, especially in public use cases.
   10.1. Mandate “Buy European” procurement rules where appropriate by amending the EU Procurement Directive (and/or through the Cloud and AI Development Act), to aid fair competition against US hyperscaler companies.
   10.2. Establish procurement requirements on ethical impact assessment, transparency, explainability, privacy, copyright compliance and meaningful civic participation depending on the AI use case.
   10.3. Design objective criteria and metrics such as the Cloud Sovereignty Framework to assess these procurement requirements while also establishing clear rules on when (certain types of) AI systems should not be procured for reasons of public accountability.

AI literacy and participation:

11. **Launch a European AI literacy strategy under the European Democracy Shield (EUDS)** to strengthen critical digital literacy, skills, and participation across population groups.
   11.1. Design tailored AI, digital, and social media literacy approaches for different age groups, socio-economic and cultural backgrounds.
   11.2. Fund civil society organisations and communities on AI literacy and participation while also monitoring these support functions performed by civil society to synchronise resources.

12. **Develop and pilot AI literacy curricula for schools and universities** through pan-European collaboration.
   12.1. Align and update guidelines for teachers and educators on disinformation and digital

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

literacy as part of the curricula, in line with the EUDS.

12.2. Couple AI literacy efforts with support for open-source and citizen coding initiatives to democratise both AI technology and its understanding.

13. **Implement the actions within the European Democracy Shield to counter misinformation** and disinformation and safeguard the integrity of the information space.

13.1. Establish the European Centre for Democratic Resilience to build capacity to anticipate, monitor and respond to information manipulation and disinformation campaigns as well as how AI can undermine democratic participation internally within the EU.

13.2. Support citizen participation, democratic innovations and civic tech in a whole-of-society approach to make societies more resilient to changing information environments, rather than merely countering foreign information manipulation and interference (FIMI).

13.3. Enhance the transparency of recommender systems, systemic risks of disinformation and labelling of AI-generated content, aligned with the DSA and AI Act.

14. **Ensure synergies between the digital regulation (AI Act, DSA) and democracy policies,** including the EUDS, European Media Freedom Act, Digital Fairness Act, and Transparency and Targeting of Political Advertising (TTPA) regulation in safeguarding free and fair elections from risks of AI.

14.1. Enhance preparedness against electoral interference on social media platforms and establish rules on fair and transparent use of AI in electoral processes, as part of EUDS.

14.2. Leverage the Digital Fairness Act to tackle undemocratic addictive design, micro-profiling and privacy violations, especially in electoral contexts.

14.3. Enforce TTPA so that AI-driven political advertising, even if disguised as entertainment or news, is properly identified and labeled.

15. **Support AI localism through a distributed, multi-level governance approach**, so that cities and local communities can play an active role in addressing governance gaps left from the EU-level.

15.1. Accelerate the emergence of local and regional AI governance models, e.g., through city AI registers that allow municipal authorities to pilot citizen-led oversight and audits of AI systems with participation of local residents.

15.2. Establish cross-directorate social and wellbeing strategies for AI that enable place-based AI responses by local and regional governments, informed by participatory policy design, foregrounding health, public services and environmental impacts.

Research and standards:

16. **Support experimental and novel research on emerging technologies that extends beyond current LLM paradigm** through DARPA-like mission-driven innovation policy under the FP10, Digital Europe programme and European Competitiveness Fund.

16.1. Prioritise explainable, low-carbon and efficient AI models by design, avoiding opaque and environmentally intensive approaches in order to address flaws in contemporary AI paradigm, e.g., in terms of world modeling and causal reasoning.

16.2. Embedded the democratic mechanisms and priorities to overarching policy and funding

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

instruments such as the Digital Decade, Digital Europe and Horizon Europe programmes, for instance in setting up societal challenge-driven AI innovation programs.

17. **Fund cutting-edge interdisciplinary research and development on democratic and responsible AI** to build digital sovereignty through cross-national measures like the European Frontier AI Initiative.
    17.1. Increase support for research that integrates ethics, law and social sciences with technical AI development, building on existing networks of excellence like ELLIS.
    17.2. Pilot the Resource for AI Science in Europe (RAISE) programme to fund state-of-the-art research that utilises AI to drive transformative scientific breakthroughs from life sciences to humanities.

18. **Standardize AI accountability by converting broad regulatory principles into accessible, enforceable benchmarks** for privacy, copyright and transparency, fit for evaluating democratic implications of AI, building on the AI Act and GPAI CoP**.**
    18.1. Introduce understandable, consumer-facing disclosure mechanisms, such as FDA-inspired "nutrition labels" detailing data sources, capabilities, risks, and limitations of AI systems, going beyond the current model card approaches.
    18.2. Guarantee meaningful explainability by requiring ethical design that incorporates sufficient friction for user awareness and oversight, rather than overly seamless AI systems.

In conclusion, democratic AI governance requires holistic efforts across multiple policy areas to encompass the entire technological lifecycle, ranging from infrastructure to deployment of AI. Overall, these recommendations seek to ensure that different dimensions of democracy like transparency and participation are not merely theoretical ideals but operational realities. However, this transition faces a tension: as the EU strengthens its sovereignty agenda, it needs to simultaneously push back on the securitization of digital policy. If sovereignty is used merely to legitimize deregulation or to replace foreign hyperscaler companies with domestic ones, Europe will only undermine the trustworthy and democratic values it seeks to protect. In other words, digital sovereignty should not come at the price of rights-based governance. This can be enabled through initiatives like the EuroStack, founded on open, auditable, and public infrastructure. By upholding fundamental rights alongside infrastructure investments, these recommendations showcase how Europe can transcend the US-China binary and position itself as a model for public-interest innovation – one that walks the talk.

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

6. Conclusion

This report represents the KT4D project's framework for democratic AI governance. It outlines the values and key approaches underpinning AI governance in the recent academic and policy discourses, covering the EU, national and international level. It identifies the key pillars of democracy – participation, freedom, equality, knowledge, transparency and rule of law – based on academic literature. In order to establish links between the pillars of democracy and the current AI governance measures, the analysis employs the AI lifecycle view. The aim is to elucidate how different governance approaches across the temporal dimension of the AI lifecycle relate to key democratic values and potentially counter power concentration. This final version of the report also includes guidance on how to translate this framework into concrete organisational practices under the AI Act. Moreover, it includes a policy roadmap and recommendations that provide practical guidance for European policymakers on how to realise the framework and govern AI systems in support of democracy.

While many of the democratic pillars are most affected in the deployment stage of AI systems, democratic considerations should cover the whole lifecycle of systems. Otherwise, influential design and development choices by private companies largely dictate how AI systems will be used, limiting the effectiveness of democratic oversight. Reflecting on the 'inverse risk pyramid' introduced in the KT4D policy brief on *Culture's Role in Navigating Technological Change* (Ahern et al., 2024), this framework similarly indicates that the problem AI presents for democracy might be less about the direct electoral manipulation, but rather about the diminishing epistemic agency of citizens and other structural risks that emerge indirectly through increasing reliance on AI systems. Because such issues are fundamentally cultural issues, the regulatory instruments can only tackle them to a limited extent. In essence, it could be that mundane spam filters and recommender systems might be more problematic than specific high-risk manipulative systems because they shape democratic practices more fundamentally, often in an inconspicuous manner.

This gives credence to a holistic, infrastructural view of democratic AI governance that is heavily interlinked with the EU's digital sovereignty. Such perspective is emphasised in the roadmap, organized around distinct policy categories, like regulatory enforcement, investments and AI literacy across three timeframes. The recommendations, on the other hand, include even more concrete measures that should be taken in the next three years to ensure that Europe reaches the desirable milestones needed for AI to reinforce democracy by 2035. The aim of the work is to outline concrete actions for EU-level policy actors, while also aiding others interested in resilient, democratic AI governance.

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

References

No	Description/Link
R1	Ahern, J., Edmond, J., Lima, E., Clarke, E., & Osipov, A. (2024). *Culture's role in navigating technological change: The KT4D perspective on recent developments in European AI policy*. Knowledge Technologies for Democracy (KT4D) Project. https://behorizon.org/wp-content/uploads/2024/03/Cultures-Role-in-Navigating-Technological-Change-The-KT4D-perspective-on-recent-developments-in-European-AI-policy-PolicyBrief-Feb2024.pdf
R2	Anderljung, M., Barnhart, J., Korinek, A., Leung, J., O'Keefe, C., Whittlestone, J., Avin, S., Brundage, M., Bullock, J., Cass-Beggs, D., Chang, B., Collins, T., Fist, T., Hadfield, G., Hayes, A., Ho, L., Hooker, S., Horvitz, E., Kolt, N., Schuett, J., Shavit, Y., Siddarth, D., Trager, R., & Wolf, K. (2023). *Frontier AI regulation: Managing emerging risks to public safety*. arXiv. https://doi.org/10.48550/arXiv.2307.03718
R3	Ashraf, C. (2020). Artificial intelligence and the rights to assembly and association. Journal of Cyber Policy, 5(2), 163-179.
R4	Bakiner, O. (2023). Pluralistic sociotechnical imaginaries in Artificial Intelligence (AI) law: the case of the European Union's AI Act. *Law, Innovation and Technology, 15*(2), 558-582.
R5	Bengio, Y. et al. (2025). International AI Safety Report. UK Department for Science, Innovation and Technology, DSIT 2025/001.
R6	Bennett, C. J., & Raab, C. D. (2020). Revisiting the governance of privacy: Contemporary policy instruments in global perspective. *Regulation & Governance*, 14(3), 447-464. https://doi.org/10.1111/rego.12222.
R7	Berlin, I. (1958). *Two concepts of liberty*. Clarendon Press.
R8	Blagovesta, N. (2024). From responsibility to risk: ethics in the Bermuda Triangle of EU research and innovation policy. *Science and Public Policy, Volume 51, Issue 2*, 207–217. https://doi.org/10.1093/scipol/scad066
R9	Bilad, M. R., Yaqin, L. N., & Zubaidah, S. (2023). Recent Progress in the Use of Artificial Intelligence Tools in Education. *Jurnal Penelitian Dan Pengkajian Ilmu Pendidikan: E-Saintika, 7*(3), 279–314. https://doi.org/10.36312/esaintika.v7i3.1377
R10	Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., ... Liang, P. (2021). *On the opportunities and risks of foundation models*. arXiv. https://arxiv.org/abs/2108.07258
R11	Bradford, A. (2022). *The Brussels effect: How the European Union rules the world* (Updated ed.). Oxford University Press. https://doi.org/10.1093/oso/9780190088583.001.0001
R12	Bradford, A. (2024). The False Choice between Digital Regulation and Innovation. *Northwestern University Law Review* 119: 377-452
R13	Buhmann, A., & Fieseler, C. (2023). Deep learning meets deep democracy: Deliberative governance and responsible innovation in artificial intelligence. *Business Ethics Quarterly, 33*(1), 146–179. https://doi.org/10.1017/beq.2021.42
R14	Bächtiger, A., Dryzek, J. S., Mansbridge, J., & Warren, M. E. (Eds.). (2018). *The Oxford handbook of deliberative democracy*. Oxford University Press.
R15	Chun, J., de Witt, C. S., & Elkins, K. (2024). Comparative Global AI Regulation: Policy Perspectives from the EU, China, and the US. arXiv preprint arXiv:2410.21279.
R16	Coeckelbergh, M. (2022). *The political philosophy of AI: An introduction*. Polity Press.
R17	Coeckelbergh, M. (2023). Democracy, epistemic agency, and AI: political epistemology in times of artificial intelligence. *AI and Ethics, 3*(4), 1341-1350.

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

R18	Cohen, G. A. (2009). *Why not socialism?* Princeton University Press.
R19	Combaz, A., Mas, D., Sanders, N., & Victor, M. (2024). *Applications of Artificial Intelligence Tools to Enhance Legislative Engagement.* AI4Democracy, IE Center for the Governance of Change.
R20	CoE, Council of Europe. (2024). The Framework Convention on Artificial Intelligence. https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence
R21	Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
R22	Dahl, R. A. (1971). *Polyarchy: Participation and opposition*. Yale University Press.
R23	de Fine Licht, K. & de Fine Licht, J. (2020). Artificial intelligence, transparency, and public decision-making. *AI & Society, 35,* 917–926. https://doi.org/10.1007/s00146-020-00960-w
R24	Dennis, C., Clare, S., Hawkins, R., Simpson, M., Behrens, E., Diebold, G., Kara, Z., Wang, R., Trager, R., Maas, M., Kolt, N., Anderljung, M., Pilz, K., Reuel, A., Murray, M., Heim, L., & Ziosi, M. (2024). *What should be internationalised in AI governance?* Oxford Martin AI Governance Initiative. https://www.oxfordmartin.ox.ac.uk/publications/what-should-be-internationalised-in-ai-governance
R25	Desai, D. R., & Riedl, M. (2024). *Between copyright and computer science: The law and ethics of generative AI*. Georgia Tech Scheller College of Business Research Paper No. 4735776. http://dx.doi.org/10.2139/ssrn.4735776
R26	de Vries, S., Kanevskaia, O., & de Jager, R. (2023). *Internal Market 3.0: The Old 'New Approach' for Harmonising AI Regulation*. *European Papers, 8*(2), 583–610. https://doi.org/10.15166/2499-8249/677
R27	Donoghue, R., Huanxin, L., Moore, P., & Ernst, E. (2024). AI, regulation, and the world of work: the competing approaches of the US and China. *Handbook on Public Policy and Artificial Intelligence,* 353-365.
R28	Du, J. (2024). The impact of artificial intelligence adoption on employee unemployment: A multifaceted relationship. *International Journal of Social Sciences and Public Administration, 2*(3), 321-327.
R29	Dung, L. (2023). Current cases of AI misalignment and their implications for future risks. *Synthese 202*, 138. https://doi.org/10.1007/s11229-023-04367-0
R30	Draghi, M. (2024). The future of European competitiveness: A competitiveness strategy for Europe. European Commission.
R31	Dworkin, R. (1986). *Law's empire*. Harvard University Press.
R32	Eaves, D., Mazzucato, M. & Vasconcellos, B. (2024). *Digital public infrastructure and public value: What is 'public' about DPI?* UCL Institute for Innovation and Public Purpose. https://discovery.ucl.ac.uk/id/eprint/10196645/1/Eaves_iipp_wp_2024-05.pdf
R33	Estlund, D. (2008). *Democratic authority: A philosophical framework*. Princeton University Press.
R34	European Commission. (2022). European Declaration on Digital Rights and Principles. https://digital-strategy.ec.europa.eu/en/library/european-declaration-digital-rights-and-principles
R35	European Commission. (2024a). A Europe fit for the digital age. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age_en
	European Commission. (2024b). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). http://data.europa.eu/eli/reg/2024/1689/oj

Funded by the European Union

KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

R36	European Commission. (2024c). First meeting of the International Network of AI Safety Institutes. https://digital-strategy.ec.europa.eu/en/news/first-meeting-international-network-ai-safety-institutes
R37	European Commission. (2024d). Over a hundred companies sign EU AI Pact pledges to drive trustworthy and safe AI development. https://ec.europa.eu/commission/presscorner/detail/en/ip_24_4864
R38	European Commission. (2025a). A simpler and faster Europe: Communication on implementation and simplification. COM(2025) 47.
R39	European Commission. (2025b). European Democracy Shield: Empowering strong and resilient democracies. JOIN/2025/791 final. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=JOIN:2025:791:FIN
R40	European Commission. (2025c). Framework for State Aid measures to support the Clean Industrial Deal (Clean Industrial Deal State Aid Framework). C/2025/3602.
R41	European Commission. (2025d). Commission work programme 2026: Europe's independence moment. COM(2025) 870 final.
R42	European Commission. (2025e). Proposal for amending Regulations (EU) 2016/679, (EU) 2018/1724, (EU) 2018/1725, (EU) 2023/2854 and Directives 2002/58/EC, (EU) 2022/2555 and (EU) 2022/2557 as regards the simplification of the digital legislative framework, and repealing Regulations (EU) 2018/1807, (EU) 2019/1150, (EU) 2022/868, and Directive (EU) 2019/1024 (Digital Omnibus). COM(2025) 837.
R43	European Commission. (2025f). 'Proposal 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)'. COM(2025) 836.
R44	European Commission. (2025g). General-Purpose AI Code of Practice (Voluntary Code of Practice for providers of general-purpose AI models). https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai
R45	Ferretti, T. (2024). Value Alignment Without Institutional Change Cannot Prevent the Societal Risks of Artificial Intelligence. LSE Public Policy Review, 3(3), 2. https://doi.org/10.31389/lseppr.113
R46	Fung, A., Graham, M., & Weil, D. (2007). *Full disclosure: The perils and promise of transparency*. Cambridge University Press.
R47	Gray, M., Samala, R., Liu, Q., Skiles, D., Xu, J., Tong, W., & Wu, L. (2023). Measurement and mitigation of bias in artificial intelligence: A narrative literature review for regulatory science. *Clinical Pharmacology & Therapeutics, Vol. 15*(4). https://doi.org/10.1002/cpt.3117
R48	Hacker, P. (2024). Sustainable AI Regulation. Common Market Law Review 61: 345-386
R49	Habermas, J. (1996). Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy (W. Rehg, Trans.). MIT Press.
R50	Farrell, H., & Han, H. (2025, August 1). AI and democratic publics: Bringing politics back into debates about AI and democracy. Knight First Amendment Institute, Columbia University. https://knightcolumbia.org/content/ai-and-democratic-publics.
R51	Helberger, N., van Drunen, M., Eskens, S., Bastian, M., & Möller, J. (2020). A freedom of expression perspective on AI in the media – with a special focus on editorial decision making on social media platforms and in the news media. *European Journal of Law and Technology, 11*(3). https://ejlt.org/index.php/ejlt/article/view/752
R52	Held, D. (2006). *Models of democracy* (3rd ed.). Stanford University Press.
R53	Howard, P. N. (2010). *The digital origins of dictatorship and democracy: Information technology and political Islam*. Oxford University Press.

Funded by the European Union

KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

R54	Ippolito, D., Tramèr, F., Nasr, M., Zhang, C., Jagielski, M., Lee, K., Choquette-Choo, C. A., & Carlini, N. (2023). *Preventing verbatim memorization in language models gives a false sense of privacy*. arXiv. https://arxiv.org/abs/2210.17546
R55	Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). *Data governance: Organizing data for trustworthy Artificial Intelligence*. *Government Information Quarterly, 37*(3), 101493. https://doi.org/10.1016/j.giq.2020.101493
R56	Jungherr, A. (2023). Artificial Intelligence and Democracy: A Conceptual Framework. *Social Media + Society, 9*(3). https://doi.org/10.1177/20563051231186353
R57	Jungherr, A., & Schroeder, R. (2023). Artificial intelligence and the public arena. *Communication Theory, 33*(2-3), 164–173. https://doi.org/10.1093/ct/qtad006
R58	Kaltheuner, F., Saari, L., Kak, A. & Myers West, S., (eds). (2024). Redirecting Europe's AI Industrial Policy: From *Competitiveness to Public Interest*, AI Now Institute.
R59	Kaminski, M. E. (2022). *Regulating the risks of AI*. *Boston University Law Review, 103*(Forthcoming). University of Colorado Law Legal Studies Research Paper No. 22-21. SSRN. https://doi.org/10.2139/ssrn.4195066
R60	Kembery, E., Bucknall, B., & Simpson, M. (2024). *Position paper: Model access should be a key concern in AI governance*. arXiv. https://arxiv.org/abs/2412.00836
R61	Khan, L. (2025). Remarks at 2025 Anti-Monopoly Summit. Available at SSRN: http://dx.doi.org/10.2139/ssrn.5491286.
R62	Kordzadeh, N., & Ghasemaghaei, M. (2021). Algorithmic bias: review, synthesis, and future research directions. *European Journal of Information Systems, 31*(3), 388–409. https://doi.org/10.1080/0960085X.2021.1927212
R63	Kreiss, D. (2016). *Prototype politics: Technology-intensive campaigning and the data of democracy*. Oxford University Press.
R64	Köbis, N., Starke, C. & Rahwan, I. (2022). The promise and perils of using artificial intelligence to fight corruption. *Nat Mach Intell 4*, 418–424. https://doi.org/10.1038/s42256-022-00489-1
R65	Lavorgna, A. (2024). Sociotechnical imaginaries of artificial intelligence in EU law making: a focus on crime and security. Im@ go. A Journal of the Social Imaginary, (24), 169-191.
R66	NIST National Institute of Standards and Technology. (2023). Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1). U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI.100-1
R67	Marshall, T. H. (1950). *Citizenship and social class and other essays*. Cambridge University Press.
R68	McKinney, S., (2024). Integrating Artificial Intelligence into Citizens' Assemblies: Benefits, Concerns and Future Pathways, Journal of Deliberative Democracy 20(1). doi: https://doi.org/10.16997/jdd.1556.
R69	Mügge, D. (2023). The securitization of the EU's digital tech regulation. Journal of European Public Policy, 30(7), 1431-1446.
R70	Narechania, T. N., & Sitaraman, G. (2024). *An antimonopoly approach to governing artificial intelligence*. *Yale Law & Policy Review*. Advance online publication. https://doi.org/10.2139/ssrn.4597080
R71	Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.
R72	Noorman, M., & Swierstra, T. (2023). Democratizing AI from a sociotechnical perspective. *Minds and Machines*. 33(4), 563-586. https://doi.org/10.1007/s11023-023-09651-z
R73	Lazar, S. (2024). Automatic Authorities: Power and AI. arXiv:2404.05990.
R74	Lazar, S. (2025). Governing the Algorithmic City. *Philosophy & Public Affairs*, 53: 102–168.

Funded by the European Union

KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

R75	Letta, E. (2024). Much more than a market: Empowering the Single Market to deliver a sustainable future and prosperity for all EU citizens. European Council.
R76	Leonelli, G. (2025). Critical Raw Materials, the Net-Zero Transition and the 'Securitization' of the Trade and Climate Change Nexus: Pinpointing Environmental Risks and Charting a New Path for Transnational Decarbonization. *World Trade Review,* 24(2), 237-256.
R77	OECD. (2019). Recommendation of the Council on Artificial Intelligence. OECD/LEGAL/0449.
R78	Ojanen, A., Björk, A. & Mikkonen, J. (2022). Promoting equality in the use of Artificial Intelligence – an assessment framework for non-discriminatory AI. Policy Brief 2022/25. Government's analysis, assessment and research activities.
R79	O'Keefe, C., Cihon, P., Garfinkel, B., Flynn, C., Leung, J., & Dafoe, A. (2020). *The Windfall Clause: Distributing the benefits of AI for the common good*. *AI Ethics and Society Conference Proceedings*. https://doi.org/10.1145/3375627.3375842
R80	Parthasarathy, A., Phalnikar, A., Jauhar, A., Somayajula, D., Krishnan, G. S., & Ravindran, B. (2024). *Participatory approaches in AI development and governance: A principled approach*. arXiv. https://doi.org/10.48550/arXiv.2407.13100
R81	Paul, R. (2022). *The politics of regulating artificial intelligence technologies: A competition state perspective*. SSRN. https://doi.org/10.2139/ssrn.4272867
R82	Pettit, P. (1997). *Republicanism: A theory of freedom and government*. Oxford University Press.
R83	Rawls, J. (1971). *A theory of justice*. Harvard University Press.
R84	Robinson, E. (2023, July 14). Contradictions abound in the EU's Critical Raw Materials Act. Land and Climate Review.
R85	Rotenberg, M. (2025). Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (Council Eur.). International Legal Materials, 64(3), 859–902. doi:10.1017/ilm.2025.1.
R86	Sastry, G., Heim, L., Belfield, H., Anderljung, M., Brundage, M., Hazell, J., O'Keefe, C., Hadfield, G. K., Ngo, R., Pilz, K., Gor, G., Bluemke, E., Shoker, S., Egan, J., Trager, R. F., Avin, S., Weller, A., Bengio, Y., & Coyle, D. (2024). *Computing power and the governance of artificial intelligence*. arXiv. https://arxiv.org/abs/2402.08797
R87	Schuett, J., Anderljung, M., Carlier, A., Koessler, L., & Garfinkel, B. (2024). *From principles to rules: A regulatory approach for frontier AI*. arXiv. https://arxiv.org/abs/2407.07300
R88	Schwabe, D., Becker, K., Seyferth, M., et al. (2024). The METRIC-framework for assessing data quality for trustworthy AI in medicine: A systematic review. *npj Digital Medicine, 7*, Article 203. https://doi.org/10.1038/s41746-024-01196-4
R89	Seger, E., Dreksler, N., Moulange, R., Dardaman, E., Schuett, J., Wei, K., et al. (2023a). Open-sourcing highly capable foundation models: An evaluation of risks, benefits, and alternative methods for pursuing open-source objectives. arXiv preprint. https://arxiv.org/abs/2311.09227
R90	Seger, E., Ovadya, A., Garfinkel, B., Siddarth, D., & Dafoe, A. (2023b). Democratising AI: Multiple meanings, goals, and methods. arXiv. https://arxiv.org/abs/2303.12642
R91	Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In proceedings of the 2019 Conference on Fairness, Accountability, and Transparency (pp. 59-68). https://doi.org/10.1145/3287560.3287598
R92	Shevlane, T., Farquhar, S., Garfinkel, B., Phuong, M., Whittlestone, J., Leung, J., Kokotajlo, D., Marchal, N., Anderljung, M., Kolt, N., Ho, L., Siddarth, D., Avin, S., Hawkins, W., Kim, B., Gabriel, I., Bolina, V., Clark, J., Bengio, Y., Christiano, P., & Dafoe, A. (2023). *Model evaluation for extreme risks*. arXiv. https://arxiv.org/abs/2305.15324

Funded by the European Union

KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

R93	Shukla, A. K., & Tripathi, S. (2024). AI-generated misinformation in the election year 2024: measures of European Union. *Frontiers in Political Science, 6,* 1451601.
R94	Sigfrids, A., Leikas, J., Salo-Pöntinen, H., & Koskimies, E. (2023). Human-centricity in AI governance: A systemic approach. *Frontiers in Artificial Intelligence, 6*, Article 976887. https://doi.org/10.3389/frai.2023.976887
R95	Smuha, N. A. (2021). From a "race to AI" to a "race to AI regulation": Regulatory competition for artificial intelligence. *Law, Innovation and Technology, 13*(1), 57–83. https://doi.org/10.2139/ssrn.3501410
R96	UK Government. (2023a). A pro-innovation approach to AI regulation. https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper
R97	UK Government. (2023b). Emerging processes for frontier AI safety. https://assets.publishing.service.gov.uk/media/653aabbd80884d000df71bdc/emerging-processes-frontier-ai-safety.pdf
R98	UNESCO. (2021). Recommendations on the Ethics of Artificial Intelligence. https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence
R99	UK Prime Minister's Office (2025). Memorandum of Understanding between the Government of the United States of America and the Government of the United Kingdom of Great Britain and Northern Ireland regarding the Technology Prosperity Deal. GOV.UK. https://www.gov.uk/government/news/memorandum-of-understanding-between-the-government-of-the-united-states-of-america-and-the-government-of-the-united-kingdom-of-great-britain-and-north
R100	Uuk, R., Brouwer, A., Dreksler, N., Pulignano, V., & Bommasani, R. (2024). *Effective mitigations for systemic risks from general-purpose AI*. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5021463
R101	von Eschenbach, W. J. (2021). Transparency and the black box problem: Why we do not trust AI. *Philosophy & Technology, 34*, 1607–1622. https://doi.org/10.1007/s13347-021-00477-0
R102	Warren, M. E. (2017). A problem-based approach to democratic theory. *American Political Science Review*, 111(1), 39–53. https://doi.org/10.1017/S0003055416000605
R103	The White House. (2023). Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. (Executive Order 14110). https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/
R104	The White House. (2025a). Removing barriers to American leadership in artificial intelligence (Executive Order 14179). https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/
R105	The White House. (2025b). America's AI action plan. https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf
R106	The White House. (2025c). Eliminating state law obstruction of national artificial intelligence policy (Presidential Action). https://www.whitehouse.gov/presidential-actions/2025/12/eliminating-state-law-obstruction-of-national-artificial-intelligence-policy/
R107	Whittlestone, J., & Clark, J. (2021). *Why and how governments should monitor AI development*. arXiv. https://arxiv.org/abs/2108.12427
R108	Widder, D. G., West, S., & Whittaker, M. (2023). *Open (for business): Big Tech, concentrated power, and the political economy of open AI*. *Big Data & Society, 10*(1), 20539517231177620. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4543807&

Funded by the European Union
KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

R109	Wong, J., Morgan, D., Straub, V. J., Hashem, Y., & Bright, J. (2022). *Key challenges for the participatory governance of AI in public administration*. Center for Open Science. https://www.turing.ac.uk/sites/default/files/2022-12/key_challenges_for_the_participatory_governance_of_ai_in_public_administration.pdf
R110	Zheng, W., Xu, X., & Hui, P. (2024). Digital democracy at crossroads: A meta-analysis of web and AI influence on global elections. In *Companion Proceedings of the ACM Web Conference 2024 (WWW '24)* (pp. 1126–1129). Association for Computing Machinery. https://doi.org/10.1145/3589335.3652003
R111	Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.

Funded by the European Union

KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.