ENSURED

Shaping cooperation
for a world in transition

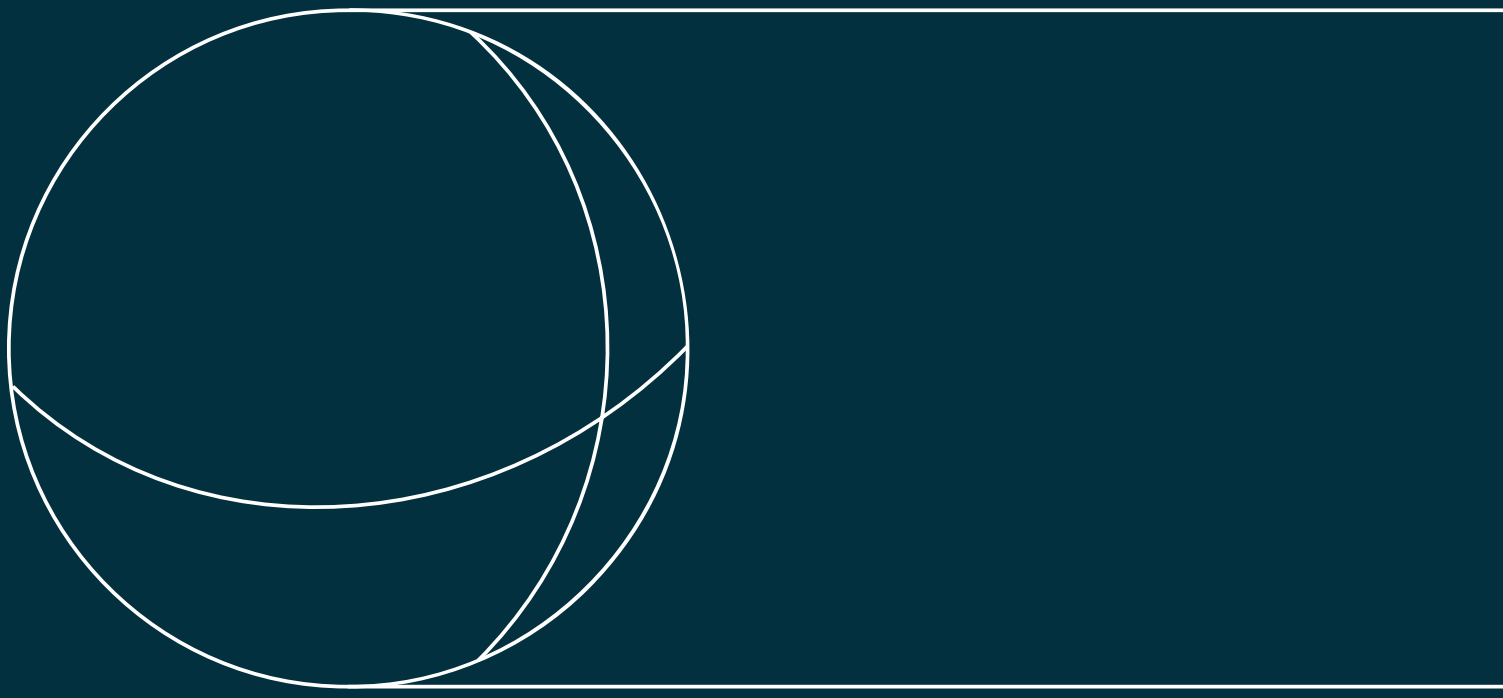
research report

Global AI Regulation at a Time of Transformation:

The Council of Europe's Framework Convention on Artificial Intelligence

Lead author: Wade Hoxtell (Global Public Policy Institute)

November 2025



Abstract

This report examines the Council of Europe's Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law, the first binding international treaty on AI, and its potential as an instrument for robust, effective, and democratic global AI regulation. While the Framework Convention's robustness is enhanced through its principles-based, flexible, and technology-neutral foundation, its effectiveness will likely be limited by exemptions of most private sector activities and national security uses of AI. Further, restricted participation in negotiations weakened democratic legitimacy, but efforts going forward hold promise due to the global openness of the treaty, and an opportunity for more participatory follow-up mechanisms. The report also highlights the European Union's central role in shaping the treaty and its potential to drive the future of the Framework Convention through implementation and its role in follow-up mechanisms.

Citation Recommendation

Hoxtell, Wade. 2025. "Global AI Regulation at a Time of Transformation: The Council of Europe's Framework Convention on Artificial Intelligence." *ENSURED Research Report*, no. 20 November 1-37. https://www.ensuredeurope.eu.

Table of Contents

Introduction

Artificial Intelligence (AI) cannot effectively be regulated by national approaches alone (Roberts et al. 2024). International coordination is necessary to manage the cross-border nature of AI markets and risks, to prevent regulatory fragmentation, and to safeguard common democratic and ethical principles. Yet, the global landscape of AI governance reveals regulatory gaps in both preventing and mitigating the potential harms of AI systems as well as in promoting safe innovation and the development of positive applications. The OECD AI Policy Observatory has documented more than 1,300 national and international policies worldwide, but the vast majority comprise non-binding frameworks rather than enforceable obligations (OECD 2025).

Further, global AI governance has become a site of contestation, reflecting wider geopolitical, economic, and normative divides. Competing approaches emphasise different values, with innovation and security on the one hand and regulation and rights on the other, while multilateral efforts struggle to bridge these divides. As a result, global AI governance has largely remained a fragmented and politically charged regime complex (Roberts et al. 2024), with no binding international agreement.

The EU needed to ensure that the Framework Convention would not conflict with the AI Act's risk-based, human-centric model.

Against this backdrop, the Council of Europe (CoE) launched a process for negotiating the first binding international treaty on AI in the spring of 2022. Although the CoE is a regional organisation, its conventions are open to accession by non-member states, allowing it to serve as a platform for developing legal standards with global reach. The negotiations brought together CoE member governments, observer states outside of Europe, the European Union, civil society, international organisations, and the private sector. Thus, this process offers insights into different actor positions on AI regulation and highlights some of the key challenges for international collaboration in this area.

After a roughly two-year negotiation period, the Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law ("Framework Convention") opened for signature in September 2024, signalling a landmark in global AI regulation. Unlike the soft-law instruments that preceded it, the Framework Convention imposes legal obligations on states that ratify it to ensure AI is developed and used in ways that respect international commitments to human rights and take democratic values into account (Council of Europe 2024a). The treaty reflects both ambition and compromise, with binding commitments for AI systems across their lifecycle, as well as trade-offs made to achieve consensus.

This report focuses on the negotiations that crafted the Framework Convention on AI and, in doing so, it provides a unique window into attempts to collaboratively govern a transformative technology. The analysis is guided by the ENSURED project's conceptual framework, which evaluates global governance institutions through three dimensions: robustness (institutional resilience and adaptability), effectiveness (capacity to deliver on goals), and democracy (inclusiveness, transparency, and

accountability) (Choi et al. 2024). Applying this lens to the Framework Convention provides a useful means to capture a snapshot of the birth of a new governance mechanism and offers a conceptual basis for assessing its potential trajectory.

This report also analyses the role of different state actors within the negotiations (focusing primarily on the European Commission and the United States, who dominated discussions), as well as civil society and private sector actors. Understanding how the EU and its member states use instruments like the Framework Convention to project standards globally is useful for evaluating both this treaty's potential impact and Europe's role in shaping global AI governance more broadly.

Its future value and impact will hinge upon whether ratifications will extend beyond Europe.

This report situates the Framework Convention within the contested landscape of global AI governance, analysing its negotiation, content, and prospects through the ENSURED framework. It finds that, while this first binding international AI treaty represents a milestone in multilateral AI governance, key compromises made in the negotiation process – including exemptions for private sector regulation and national security uses of AI – will likely weaken the Framework Convention's effectiveness. On the other hand, these compromises allowed for a more robust treaty by prioritising its global accessibility and adaptiveness to new technological developments. At the same time, the limited nature of non-state actor participation, the absence of other major global AI actors such as China, and the relatively narrow range of like-minded state actors raise questions about the democratic inclusiveness and legitimacy of the treaty. Its future value and impact will hinge upon whether ratifications will extend beyond Europe, how states will implement its principles domestically, and whether follow-up mechanisms will succeed in promoting accountability and deepening participation.

Global AI Governance and the Council of Europe

Regulating AI is challenging for a number of reasons. This report zeroes in on three of the main difficulties. First, as with past cases of new technologies, regulatory bodies are struggling to keep pace with the rapid technological advancements in AI systems, as well as the emergence of new use cases and risks (Wallach and Marchant 2019). As such, AI regulation is subject to the so-called Collingridge Dilemma. Also called "the pacing problem," the Collingridge Dilemma arises when a given technology's impacts are still uncertain, but delaying regulatory action to understand these impacts creates path dependencies that then hinder future regulation (Collingridge 1980).

Much like the emergence of the internet in the 1990s, governance mechanisms will need to react nimbly to new developments and uses of AI across the entire spectrum of society. This means confronting the political challenge of coordination among powerful – and often competing – states and corporate actors, the institutional challenge of designing adaptable yet enforceable rules, and the normative challenge of defining what constitutes responsible and legitimate use of AI across diverse political and cultural contexts.

Second, the global regulation of AI is becoming increasingly (geo-) politicised: it is impacted by both differing domestic political and economic considerations and fragmented ideological approaches. The dual-use nature of AI that includes both beneficial innovation and harmful applications comes with significant trade-offs (Bremmer and Suleyman 2023), and global actors approach these trade-offs differently, as evidenced by, for example, the AI Action Summit in Paris in 2023 as well as negotiations of the Framework Convention on Artificial Intelligence (detailed below).

The global regulation of AI is becoming increasingly (geo-)politicised.

These and other fora have seen actors and coalitions form around different conceptualisations of what regulation should look like. The US and the United Kingdom, for example, promote a 'light' approach to regulation that prioritises economic growth and national security interests, while the EU is spearheading an approach that focuses on risk management and the protection of human rights. Other states, including India, Brazil, and Canada, aim to strike a balance between innovation, deployment, and regulation (Bazoobandi et al. 2025).

States are not the only actors shaping the conversations around AI regulation. The concentration of power in large private companies represents another significant factor influencing discussions, and it triggers a number of concerns, including threatening the legitimacy of states and the ability to hold these companies accountable for the impact of their technologies (Zhang et al. 2025).

A third challenge for global AI regulation: ensuring equitable participation in the global effort to regulate such a transformative new technology.

Existing voluntary principles around AI regulation are enshrined in, for example, the Organisation for Economic Cooperation and Development's (OECD's) *AI Principles* and the United Nations Educational, Scientific and Cultural Organization's (UNESCO) *Recommendations on the Ethics of Artificial Intelligence*. These have achieved intergovernmental traction and acknowledge the importance of inclusiveness and multi-stakeholder input, but building effective mechanisms to ensure such equitable engagement remains difficult (Buhmann and Fieseler 2021).[1]

The concentration of power in large private companies represents another significant factor influencing discussions.

Yet, actors such as the US, China, and the EU dominate discussions on AI regulation, while non-state actors as well as states from the Global South are underrepresented (Roche et al. 2021). Ensuring meaningful participation for these actors goes beyond recognising different stakeholders as legitimate participants in consultations. It means equitably providing all actors access to information, giving them the opportunity to influence agendas and outcomes, and sustaining their participation over time (Taggart and Haug 2024). As examples like the AI Action Summit show, large AI actors pursuing technological leadership and economic advantage can crowd out the voices of other states and civil society actors, which complicates inclusivity in AI regulation discussions. Without meaningful participation of a wide range of public, private, and civil society actors, however, the legitimacy of and trust in global regulatory frameworks for AI will be weakened.

At the international level, these three challenges have fuelled a prevailing trend of creating non-binding, multilateral AI frameworks that emphasise voluntary principles, guidelines, codes of conduct, and standards (UNU 2023; Cole 2024). Arguably, the most prominent of these are the OECD's *Principles for Trustworthy AI*, as well as the OECD AI Policy Observatory's work, which serves as a knowledge hub on emerging AI policies across the world (OECD 2024, 2025). Intergovernmental efforts, such as the United Nations Global Digital Compact and its High-Level Advisory Body on AI, also represent this push for voluntary regulation. While they provide a platform for broader engagement between states (United Nations 2025) and are useful for providing a foundation for other regulatory efforts at the subnational, national, regional or international levels (Interview 6), they remain in the realm of soft law, carrying some normative weight, but lacking enforcement mechanisms.

This backdrop illustrates the significant challenge of creating internationally binding obligations for AI that are clear, specific, compatible with existing national legal systems, and politically palatable in different national contexts. The Council of Europe's Framework Convention represents an attempt to tackle this challenge by creating the first legally binding treaty on AI with obligations, as well as a review mechanism for tracking compliance.

1 The OECD AI Principles have been adopted by all 38 OECD member states, 10 non-member states, as well as the EU (OECD 2024). The Recommendations on the Ethics of AI were adopted in 2021 and are applicable to all 194 members of UNESCO (UNESCO 2023).

The Council of Europe and Initial Steps Toward a Binding Treaty on AI

The CoE was founded in 1949 to uphold human rights, democracy, and the rule of law across its now 46 member states. While the CoE's primary and most well-known role is stewarding the European Convention on Human Rights (ECHR), it has also developed other influential treaties, such as the Budapest Convention on Cybercrime and the Convention 108+ on data protection. As such, the CoE has shepherded digital governance mechanisms that also gained traction beyond European borders.

In the mid-2010s, the CoE began examining the societal impacts of AI and concluded that some form of regulation would be beneficial (Interview 2). In July 2019, its Committee of Ministers formally identified AI as a priority issue, stressing that the CoE could add value to AI regulation efforts due to its unique, continent-wide legal space where member states have a legal obligation to guarantee the same rules on human rights, democracy, and the rule of law (Council of Europe 2019b). Further, the Committee highlighted the global reach of the CoE's existing conventions, as well as its experience in leveraging multi-stakeholder cooperation processes with civil society, the private sector, and academia (Council of Europe 2019b).

The CoE has shepherded digital governance mechanisms that also gained traction beyond European borders.

In September 2019, the CoE launched the Ad Hoc Committee on Artificial Intelligence (CAHAI) (Council of Europe 2019a) to advise on the feasibility and design of a potential legal framework for governing AI (Schneider 2024; Council of Europe 2019b). The CAHAI comprised CoE member state representatives, non-member observer states (e.g., US, Canada, Japan), intergovernmental organisations (e.g., EU, OECD, and UN bodies), and non-state stakeholders (civil society, academia, private sector). Formal decision-making power rested with CoE member states, while observers and non-state actors could intervene, provide written input, and participate in working groups (Council of Europe 2019a).

In December 2020, the CAHAI published two reports. The first, *Towards Regulation of AI Systems*, mapped AI's impacts on human rights, democracy, and the rule of law, featured a survey of existing soft law and ethical guidelines, and three case studies of national efforts in Israel, Japan, and Mexico. Notably, the report concluded that a binding instrument would be a useful complement to existing non-binding instruments (Council of Europe 2020b). The second report was a feasibility study that laid the groundwork for the CoE's process moving forward, accounting also for the CoE's existing standards and commitments on human rights, democracy, and the rule of law, as well as equivalent standards elsewhere (Schneider 2024). Like the first report, this study made the case for creating a new, legally binding Framework Convention, rather than opting to establish a non-binding instrument or to modernise existing binding instruments (Council of Europe 2020a).

This binding option found broad support when the CAHAI launched a multi-stakeholder consultation with over 200 participants to discuss a possible

AI governance mechanism in May of 2021. There were, however, some diverging views on whether to directly regulate the private sector, as well as on how strong an eventual monitoring and enforcement mechanism should be (Council of Europe 2021). A smaller number of respondents, particularly from the private sector and technical communities, expressed concern that overly rigid or uniform rules might stifle innovation or prove unable to keep pace with technological change. Others favoured a more incremental or sector-specific approach, or argued that voluntary self-regulation could complement or even substitute binding measures (Council of Europe 2021).

Building on the CAHAI's feasibility study as well as a primer published by the Alan Turing Institute in June of 2021, the CAHAI released its key report in December of the same year: *Possible Elements of a Legal Framework on Artificial Intelligence*. This report outlined the building blocks that would eventually form the final Framework Convention.

Negotiating the Framework Convention on AI: Positions of Key Actors

With the conclusion of the CAHAI's mandate at the end of 2021, the Committee of Ministers established a new Committee on Artificial Intelligence (CAI) and mandated the CAI to negotiate and draft a binding treaty (Council of Europe 2022; Schneider 2024). The CAI was structured similarly to the CAHAI, namely as an intergovernmental negotiating body, composed of representatives from all 46 CoE member states with voting rights, alongside observer states, representatives from intergovernmental organisations, civil society, the private sector, and other experts (Council of Europe 2022).

The negotiations were dominated by the European Commission and the US; both actors set much of the agenda and shaped the key provisions of the treaty (Interviews). For the Commission, the process was intertwined with the EU's discussions around its own domestic regulatory project, the AI Act. The EU needed to ensure that the Framework Convention would not conflict with the AI Act's risk-based, human-centric model while, at the same time, it wanted to inject its regulatory philosophy into the Framework Convention (Interview 4). Given that the EU's legislative process around the AI Act was still in a critical phase in 2022, Commission officials purposefully slowed the CAI negotiations until they finally secured a formal mandate and received negotiation guidelines from the EU member states in November 2022 (Interviews 3 and 4). This meant that negotiations on the Framework Convention did not begin in earnest until spring of 2023.

For US officials, the Framework Convention was an opportunity to ensure that the world's first binding international treaty on AI was aligned with US interests.

For US officials, the Framework Convention was not only an opportunity to ensure that the world's first binding international treaty on AI was aligned with US interests, but it also represented a means for finding agreement between like-minded democracies on what democratic states should and should not do with regard to using AI (Interview 4). The Biden administration believed that a binding Framework Convention could create consensus around responsible, rights-based government use of AI (Interview 1) — in that way, the treaty could feature as a tool in the geopolitical contest with rivals such as China (Interview 2). However, the sticking point for US negotiators was how to shape the treaty such that it was meaningful, while avoiding getting ahead of domestic policymaking and protecting flexibility on some key aspects, such as the direct regulation of the private sector (Interviews 1 and 4). With the US Congress still debating its domestic approach to potential AI regulation and the Biden administration limited to policy tools within the bounds of existing US law (such as executive orders), US officials lacked a comprehensive framework to guide their negotiating positions. Given this, US officials supported the development of high-level principles and pushed a narrowed focus

of regulating only public uses of AI, which they saw as the most feasible basis for consensus (Interviews 1 and 4).

Depth vs. Breadth

From the start, it was clear that the Framework Convention would be binding, but a key tension in the negotiations was the trade-off between the *depth* of the treaty (how precise and prescriptive its provisions would be) and its *breadth* (how attractive and accessible it would be to a wide range of potential signatories). Many EU member states, including Germany, France, Spain, the Czech Republic, Estonia, Ireland, Hungary, and Romania, as well as the European Commission, emphasised that securing broad participation was more important than drafting an ambitious text that would only get limited global uptake (Murphy and Jacobson 2024; Bertuzzi 2024; Interview 6).

A key tension in the negotiations was the trade-off between the depth of the treaty and its breadth.

Notably, EU member states in any case would be subjected to the emerging comprehensive regulatory framework of the AI Act, and EU officials were not only aware that compliance with the AI Act would satisfy their obligations under the Framework Convention, but it guided their entire negotiation posture (Interviews 1, 2, and 6). Given that the Framework Convention would not weaken rules already put into place through the AI Act, European negotiators conceded that a more flexible and less prescriptive text for the Framework Convention, which still carried forward a rights-based vision, was the price of achieving a more global instrument (Interview 4). The US, backed by Canada, Japan, and the UK, also pressed for breadth and flexibility, especially in relation to private sector obligations, which would become one of the main sticking points in the negotiations (Interviews; Volpicelli 2024).

Civil society organisations (CSOs), such as AlgorithmWatch, Access Now, and ECNL, consistently warned that foregoing enforceable rights risked diluting the treaty into a declaration of principles with little binding force (Access Now 2023). They pressed for stricter obligations on, for example, the primacy of international human rights law over national discretion, independent oversight mechanisms, and explicit prohibitions on unacceptable AI uses (Access Now 2023).

Despite this input, negotiations swung in the direction of prioritising breadth over depth due to significant support amongst government officials and the preference of the CAI to make the treaty flexible and attractive to a number of signatories (Interviews 4 and 6).[2] However, CSOs did secure some wins to give the treaty additional depth, such as a requirement in Article 16 for signatories to conduct human rights, democracy, and rule of law impact assessments, as well as some additional language in the accompanying

2 As Thomas Schneider, Chair of the CAI, argued, "effective AI governance had to be binding yet flexible, anchored in shared values like human rights and democracy, but designed in a way that gave potential signatories room to adapt obligations to their domestic legal frameworks" (Schneider 2024).

Explanatory Report, for example, on the importance of environmental impact assessments (Interview 3; Council of Europe 2024b).[3]

Another minor area of contention concerned the extent to which the treaty should cover the early stages of AI research and development. Civil society groups pressed for clear obligations extending across the entire AI lifecycle, arguing that excluding early-stage research and development (R&D) would weaken accountability and allow harmful systems to be developed without scrutiny (Access Now 2023). Some CSO representatives also called for mandatory human rights and democratic impact assessments at the design stage, particularly for high-risk or public sector systems (Interview 3). However, government officials broadly resisted prescriptive rules on research, fearing that they would stifle innovation. Instead, they emphasised that the Framework Convention should focus on the use and deployment of AI, when and where human rights and rule-of-law risks actually materialise (Interview 4).

Civil society groups pressed for clear obligations extending across the entire AI lifecycle.

The compromise was to craft language that stressed a life-cycle approach and acknowledged the risks that occur during the development and design stages of AI systems, while leaving implementation flexible and non-prescriptive (Council of Europe 2024a; Schneider 2024). While debate around R&D provisions proved to be minor compared to the sharper conflicts over private sector regulation and national security, it revealed the widespread reluctance amongst government officials to extend binding obligations beyond existing policies on AI research and development, for example, those set out in the AI Act in the case of the EU.

Participation vs. Expediency

The negotiations were dominated by governments; non-state participation by civil society, experts, and private sector actors was limited during official proceedings. One reason for this was due to CoE rules that only permit actors with official observer status to take part in negotiations (Interview 7). Any actor not already recognised as a member or observer requires formal admission by the CoE, a process that involves a review and full consensus of member states. This structural rule likely prevented greater representation and participation of civil society and private actors in the negotiation process (Interviews 3 and 7).

Those civil society and private sector representatives who were present had limited influence over the final result. First, non-state actors were excluded from the drafting group that shaped the treaty text (Interviews). In addition, documents coming out of the drafting group were classified, which restricted non-state actors' ability to gather input from their respective members or supporters and provide comments (Interview 7). Government officials justified their support for a states-only drafting group as necessary for expediency in the negotiations. It was also meant to encourage more candid state-to-state discussions, since information

3 The Explanatory Report is worth reading as it highlights key aspects of the treaty and their intent.

and positions had been leaking, complicating negotiations (Interview 6; Bertuzzi 2023). The CAI did circulate treaty drafts to non-state participants to comment on, but given the negotiations' ambitious timeline, they had little time to do so meaningfully. CSOs often had to restate their comments on the treaty drafts during the plenary meetings to be certain that they were acknowledged (Interview 3). However, government officials stressed that CSOs did influence negotiating positions with their arguments and, ultimately, impacted language in both the body of the treaty and, in particular, the accompanying Explanatory Report (Interview 1).

Private sector participants largely limited their input in negotiations.

Private sector representatives were reserved in giving input, as they realised early in the process that the eventual design of the Framework Convention would largely be uncontentious for them (Interviews 1, 2, and 7). They recognised the value of a binding international treaty, particularly one focusing on human rights, democracy, and rule of law protections (Interview 7; International Chamber of Commerce 2023), and they were largely aligned on their key positions with regard to how such a treaty should look.[4] Their negotiation position was primarily guided by their desire for coherence with existing frameworks (notably, the OECD AI Principles and the emerging EU AI Act) to avoid conflicting obligations and excessive compliance burdens. In addition, they argued for a clear and narrow definition of AI, aligned with OECD standards (Interviews 1, 2, and 7). They also advocated for a treaty that supported the spread of AI's benefits without creating restrictive, top-down rules; they emphasised differentiated responsibility across the AI lifecycle, cautioning against blanket obligations that might apply unevenly to different types of companies (Interview 7).

Given that these priorities proved to be the prevailing sentiment among state negotiators, and because it was clear early on that the treaty would be based around high-level principles to be interpreted by national governments and not create new, direct obligations for companies, private sector participants largely limited their input in negotiations. US companies anticipated that the treaty itself would not lead to new regulations on them directly, whereas EU-based companies would, anyhow, be subject to the AI Act's compliance obligations, rendering the practical implications of the treaty comparatively limited for them.

Opt-in vs. Opt-out

One of the most contentious issues in the negotiations was whether the treaty's obligations should extend to the private sector or not (Interviews). The debate centred on two competing models. The first was an opt-in approach, where private companies would be excluded from the treaty by default, and any explicit regulation of companies would rely on state

4 Private sector participation included, for example, Meta, Microsoft, IBM, IEEE, Telefonica, and the International Chamber of Commerce. Tellingly, private sector participation in plenaries dropped considerably once the consolidated working draft of treaty was released in July 2023. Lists of participants of all plenary meetings can be found here: https://www.coe.int/en/web/artificial-intelligence/cai.

parties' decision to opt into applying Framework Convention provisions to the activities of private entities. The second was the opt-out approach, which would by default obligate state parties to apply the treaty's provisions to the activities of private companies, but would allow them the option to exempt themselves from this requirement. While the opt-in model risked undermining the effectiveness of the treaty and the consistency of AI standards across countries, negotiators saw it as a way to attract more signatories, particularly those unwilling or politically unable to regulate their domestic companies (Interviews 1, 3, 4, 6, and 7). Conversely, the opt-out model promised a stronger baseline of regulation, but carried the danger of deterring key players like the US or the UK (Bertuzzi 2024).

EU officials strongly supported the opt-out model, arguing that excluding private companies outright would "undermine the treaty's effectiveness in regulating AI's impact on society" (Bertuzzi 2024). Further, Commission officials emphasised that it was crucial for the Framework Convention to send the message that both public administrations and private companies have responsibilities, even if the treaty granted flexibility on how to carry them out (Interviews 3 and 4). The opt-out model also aligned most closely with the EU AI Act, which set obligations on both public and private entities involved in AI development and deployment (Interviews 1 and 4).

The US, on the other hand, made it clear from the beginning that it could only accept binding regulations on the public use of AI, not on the private sector. The only exception to this position was the case of private companies acting on behalf of public agencies, a provision which was aligned with already existing policy and legal authority previously granted through the US Congress (Interview 1). This negotiating position reflected the domestic political realities and the limits of existing executive authorities to regulate private entities' AI activities. The Biden administration's Executive Order on AI, for example, largely applied only to federal agencies (White House 2023; Interview 1). US negotiators, therefore, insisted on the opt-in model, arguing that this was the only way Washington could sign on (Interview 1). Canada, Japan, and the UK backed this US position (Interviews).

The US made it clear from the beginning that it could only accept binding regulations on the public use of AI, not on the private sector.

Ultimately, the opt-in approach prevailed, reflecting significant US influence on the process. The text of the final treaty allows signatories to choose whether and how to apply treaty obligations to private companies,[5] but with a stipulation, successfully argued for by the European Commission, that requires them to clarify this when ratifying (Council of Europe 2024a). For the US, this solution allowed it to sign the Framework Convention while avoiding the political inconvenience of openly exempting companies from an international human rights treaty.

5 The treaty does cover, by default, private companies working on behalf of public authorities: "Each Party shall apply this Convention to the activities within the lifecycle of artificial intelligence systems undertaken by public authorities, or private actors acting on their behalf" (Council of Europe 2024a). This was mostly uncontentious for the U.S. since this provision aligned with the Biden administration's Executive Order on AI (later rescinded by the Trump Administration).

Civil society actors strongly opposed this compromise, arguing that "most risks to human rights, democracy, and rule of law originate from the way AI is designed and used by the private sector" (Caunes 2024). Leaving companies outside the scope, they argued, "would send a dangerous signal" and would risk turning the treaty into little more than symbolic law (AlgorithmWatch 2024).

National Security

The other major sticking point in the negotiations was the treatment of national security-related uses of AI (Interview 1, 3, 4, and 6). During the initial stages of negotiations, the European Commission did not have full clarity on its mandate in this area due to an ongoing debate in the context of the AI Act discussions. While the European Commission's original 2021 proposal for the AI Act excluded "military purposes," it did not explicitly exempt "national security" uses of AI. This sparked debate among member states, who argued that all security, intelligence, and defence applications must remain outside the AI Act's scope, in line with Article 4(2) of the Treaty of the European Union (Interview 4; Council of the European Union 2022; Powell 2024). As a result, the final text of the AI Act broadened the exemption to exclude all "military, defense, and national security purposes." Once this was resolved in the context of the AI Act, it also clarified the European Commission's negotiation stance in the Framework Convention process. The Commission did not receive a mandate from the Council of the European Union to include national security in the scope of the Framework Convention negotiations (Interview 4).

US officials argued in favour of including national security uses of AI in the treaty.

The US took a different view. Recognising that the Framework Convention would be open to ratification by any government, US officials argued in favour of including national security uses of AI in the treaty. The reasoning was that including the regulation of national security activities would create a situation where countries that disrespect human rights (such as China) could not become party to the treaty if they did not agree to the obligations on regulating national security uses of AI (Interview 3). The US position was that if this treaty was to become a global standard, even if only through the creation of norms, it would be important that those states that do respect human rights, democracy, and the rule of law should not make exceptions using the justification of national security purposes (Interviews 1 and 3).

The US ultimately accepted that the EU and other states could not shift their position, culminating in Article 3 of the Convention, which explicitly carves out national security from its coverage (Council of Europe 2024a). Civil society strongly criticised this carve-out, warning that national security exemptions risked becoming a blanket cover for rights-violating uses of AI (AlgorithmWatch 2024). However, civil society did secure a victory in how the drafters framed this issue in the final text, which describes the understanding that AI-related national security activities exempted by the Convention "are conducted in a manner consistent with applicable international law, including international human rights law obligations, and with respect for its democratic institutions and processes" (Council of Europe 2024a; Interview 3).

The Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law

The text of the Framework Convention on Artificial Intelligence was finalised in March 2024 and formally adopted by the CoE's Committee of Ministers in Strasbourg, France, on May 17, 2024 (Council of Europe 2024a). The treaty opened for signature in September 2024, with early signatories including Andorra, Georgia, Iceland, Israel, Moldova, Norway, San Marino, the UK, the US, and the EU on behalf of its 27 member states (Council of Europe 2024). For the treaty to enter into force, it must be ratified by a minimum of five states, including three CoE members.

Table 2: Milestones of the CoE's Framework Convention on Artificial Intelligence

Continued on the next page.

Date	Milestone
February 2019	The CoE convenes the conference "Governing the Game Changer – Impacts of artificial intelligence development on human rights, democracy and the rule of law."
May 2019	The CoE Commissioner for Human Rights releases the recommendation *Unboxing Artificial Intelligence: 10 Steps to Protect Human Rights*.
July 2019	At the 129th Session of the Committee of Ministers, the CoE identifies regulation of AI as a priority.
September 2019	The Committee of Ministers adopts terms of reference of the CAHAI.
December 2020	The CoE publishes (1) a feasibility study and (2) a report on the potential impact of AI systems on human rights, rule of law, and democracy.
March-May 2021	A multi-stakeholder consultation process gathers 260 responses with input on the main elements of an eventual new legal framework.
June 2021	The Alan Turing Institute publishes a primer on the main concepts and principles in the CAHAI's feasibility study.
December 2021	CAHAI releases *Possible Elements of a Legal Framework on Artificial Intelligence, Based on the CoE's Standards on Human Rights, Democracy and the Rule of Law* which lays out key pieces of what would eventually become the Framework Convention on Artificial Intelligence.
January 2022	CAI begins its mandate, superseding the CAHAI.
April 2022	First plenary of the CAI is held.
June 2022	CoE Council of Ministers mandates the CAI to lead a negotiation process for launching a new binding international agreement.
November 2022	Council of the European Union gives mandate to the European Commission to negotiate the Framework Convention.

Continued from the previous page.

February 2023	Revised version of the zero draft gets released to the public.
July 2023	A consolidated Working Draft of the Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law is released.
March 2024	Negotiations completed and the final draft of the Framework Convention on AI, Human Rights, democracy and the Rule of Law is released.
May 2024	Committee of Ministers adopts the Framework Convention on Artificial Intelligence.
September 2024	The Framework Convention opens for signatures. Andorra, Georgia, Iceland, Israel, Moldova, Norway, San Marino, the UK, the US, and the EU (on behalf of its 27 Member States) sign it.
November 2024	CAI publishes the HUDERIA framework.
November 2024 – September 2025	Montenegro, Lichtenstein, Canada, Japan, Switzerland, Ukraine, and Uruguay sign the treaty.
June 2025	European Commission ratification proposal is sent to the Council of the European Union.
December 2025	Mandate of the CAI ends.

The final text of the Framework Convention covers AI systems across their entire lifecycle, from design through deployment to decommissioning, regulating activities related to AI and their impacts rather than the technology itself (Interview 2). It rests on three main pillars (Council of Europe 2024a):[6]

1. **Legally binding obligations** that require state parties to apply their existing human rights, democracy, and rule of law commitments to the field of AI. These binding obligations are codified as fundamental principles (see Box 1) that AI-related activities must comply with throughout their entire lifecycle. Further, the Framework Convention adopts a graduated and differentiated approach, meaning regulatory measures must be proportionate to the specific risks and impacts posed by different AI systems, leaving flexibility regarding how to satisfy these obligations.

Box 1: Fundamental principles

- Human dignity and individual autonomy
- Equality and non-discrimination
- Respect for privacy and personal data protection
- Transparency and oversight
- Accountability and responsibility
- Reliability
- Safe innovation

6 Although not formally part of the Framework Convention, the Explanatory Report to the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law provides an interpretive aid and is worth reading. It provides interesting insight into the intentions of the drafters and, importantly, could be consulted by the European Court of Human Rights in assessing the scope and application of the Convention's provisions (Interview 2; Council of Europe 2024b).

2. **Remedies, procedural rights, and safeguards**, which must be made available to any person whose human rights and fundamental freedoms are significantly impacted by an AI system. Individuals must be able to access information about AI systems that affect them, receive notice when interacting with AI, and challenge decisions before competent authorities.

3. **Ongoing risk and impact assessments** on human rights, democracy, and the rule of law, and the **implementation of prevention and mitigation measures**. Authorities should also be allowed to impose bans or moratoria on certain AI applications.

Table 1: Milestones of the CoE's Framework Convention on Artificial Intelligence

	Robustness	Effectiveness	Democracy
European Commission	Promoted a principles-based, risk-oriented treaty, accepting flexibility to broaden participation and breadth. Saw flexibility and rights-based anchors as key to long-term durability.	Advocated binding obligations on both public and private actors. Requirement to exempt national security uses of AI reduced effectiveness.	Embedded democratic rights and remedies in the treaty, but compromised on inclusivity by limiting non-state participation in negotiations.
United States	Pursued robustness through flexibility in approaches to national implementation. Emphasised attracting more signatories through political sustainability over ambitious obligations.	Limited effectiveness by rejecting binding private sector obligations and agreeing to EC's demands for national security exemptions. Saw high-level principles as sufficient to make the treaty feasible and broadly attractive.	Agreed to limit non-state actors role in drafting, reducing inclusivity.
Civil Society	Warned that exemptions and principle-only commitments would weaken resilience. Called for strong, enforceable provisions to prevent erosion of obligations over time.	Sought enforceable bans, lifecycle impact assessments, and oversight to maximise effectiveness. Argued national security carve-out and opt-in approach private sector would leave key risks unchecked.	Criticised exclusion and lack of transparency in negotiations, and pressed for participation and accountability. Linked inclusivity to participation of civil society, substantive legitimacy and protection of rights.
Private Sector	Favoured coherence with OECD and EU frameworks, stressing clear definitions and predictability. Opposed prescriptive measures that might constrain flexibility or innovation.	Defined effectiveness as avoiding fragmentation and excessive burdens. Supported risk-based rules and coherence, but resisted heavy compliance costs or disclosure obligations.	Sought consultation and transparency to avoid fragmentation, accepting states' final authority. Framed democracy in terms of stakeholder participation rather than binding obligations.

Robustness

The robustness of the Framework Convention rests on its capacity to withstand shifting political, social, and economic contexts, rapid technological developments in the AI field, as well as on the suitability of the CoE as an institution. In this regard, the treaty negotiators attempted to increase the Framework Convention's robustness in three ways. First, it was deliberately crafted as a technology-neutral, principles-based instrument. By establishing high-level obligations on human rights, democracy, and the rule of law, the Framework Convention is intended to remain relevant, even as both technologies and political climates change, thus increasing its robustness.

The Framework Convention is intended to remain relevant, even as both technologies and political climates change, thus increasing its robustness.

Crucially, robustness is not only a technocratic matter, but also one of political durability, and cracks have already started to form. By giving states flexibility in how they implement the treaty obligations (by accommodating diverse legal systems and political contexts), the Framework Convention enhanced its robustness. Yet, the geopolitical environment for AI governance has shifted dramatically since negotiations began. Early momentum for binding international rules has been overshadowed by narratives centred on economic competitiveness and national security (Interviews 1 and 6; Hofmann and Pawlak 2024). While it seems likely that the treaty will indeed achieve enough ratifications to enter into force (despite the change in political wind), it remains to be seen whether it obtains enough ratifications to create conditions for meaningful follow-up with noticeable and sustainable impact.

Second, by framing obligations around activities throughout the AI lifecycle rather than specific aspects of AI systems themselves, the treaty ensures resilience against obsolescence and provides a stable baseline for states to adapt within their domestic systems. The treaty's ability to absorb new AI developments without needing amendments (and inevitably, political wrangling) is meant to contribute to its stability and durability. This approach mirrors earlier CoE instruments, like the Convention on Cybercrime, which has endured precisely because it was structured around rights and principles rather than tied to specific technologies (Juncher 2025).

In addition, while excluding obligations on regulating the private sector or national security uses of AI made the treaty more attractive to a greater number of potential signatories and increased its potential robustness, these decisions likely came with a trade-off to its effectiveness.

Finally, the robustness of the Framework Convention is also a function of the suitability of the CoE as host for this kind of instrument. The CoE's legitimacy in the field of AI regulation may not be obvious compared to other institutions, but its case is strengthened by its long-standing mandate, its strengths in human rights, democracy, and the rule of law, and its precedent-setting conventions on cybercrime and data protection. The Framework Convention can leverage the CoE's existing legal and procedural apparatus, including its human rights directorate and treaty bodies, to ensure coherence with cornerstone instruments such as the

European Convention on Human Rights. Further, the CoE has built up expert capacity on digital issues in different committees, and it has plans to further build on this through the extension and expansion of the mandate of the CAI (Interview 6).

Smaller member states have already used CoE resources for capacity-building and guidance.

In addition, smaller member states have already used CoE resources for capacity-building and guidance, for example, when applying the Human Rights, Democracy, and Rule of Law Impact Assessment for Artificial Intelligence (AI) Systems methodology (HUDERIA) (Interviews 2 and 5). HUDERIA, a voluntary tool developed by the CoE together with the Alan Turing Institute, helps actors assess and mitigate the risks of AI and is designed to be context-sensitive and resilient in the face of technological changes and political shifts, making it more robust as a tool. Notably, it is also available to all states, even if they have not signed the treaty (Interview 5). However, while HUDERIA contributes to robustness by enabling greater context-specific usage, this flexibility may come at the expense of effectiveness, due to the potential for inconsistent implementation of treaty provisions across jurisdictions. Further, it remains to be seen whether funding will be made available to continue with outreach and capacity-building efforts like the HUDERIA Academy (Interview 5).

Finally, the CoE's role as shepherd of the Framework Convention, and as an actor in the AI regulation space more generally, may have its limits. The organisation is less visible in regulatory debates compared to the European Commission and budgetary constraints, exacerbated by Russia's expulsion following its invasion of Ukraine, raise questions about the CoE's long-term capacity to manage the Framework Convention and push it forward (Interviews 3 and 6).

Effectiveness

The effectiveness of the Framework Convention will be determined by its ability to ensure that, within the breadth and depth of its coverage, AI systems are developed and used in ways that respect human rights, democracy, and the rule of law. Since the treaty excludes private-sector and national-security applications, its effectiveness is understood specifically in terms of its influence on public-sector governance, national implementation practices, and its capacity to shape global norms. In this context, the key elements that make or break the treaty's effectiveness relate to (1) its ambition, (2) its potential for attracting signatories, (3) national-level implementation of domestic regulation, and (4) the treaty's follow-up mechanisms.

Ambition

The Framework Convention is legally binding. Yet, the treaty does not create new rights; it applies AI-specific challenges to existing rights and obligations. On the one hand, this may reduce the treaty's effectiveness. On the other hand, this approach can contribute to robustness since international conventions and national laws around, for example, human

rights are already in place, and risks arising from new technologies can be grafted over these existing rights (e.g., right to privacy), which can make legal interpretation clearer (Interview 3).

The Framework Convention applies to all stages of the AI lifecycle, from design through deployment to decommissioning. Its focus is not the technology itself but rather its impacts, for example, whether an AI system undermines privacy, equality, human autonomy, or other fundamental rights. Concretely, this requires state parties to enshrine human dignity and autonomy as legally binding guardrails (Council of Europe 2024a). For example, a judge applying domestic law can use the Framework Convention to ensure that AI systems cannot replace or override human decision-making in ways that undermine fundamental rights (Interview 6). Unlike the EU AI Act, which is largely a market-regulation tool, the Framework Convention provides a rights-based legal foundation for the national regulation of AI.

The Framework Convention provides a rights-based legal foundation for the national regulation of AI.

While the treaty's flexible approach boosts its robustness, similar to the Budapest Convention on Cybercrime, which still remains relevant decades after its launch (Juncher 2025), it is less clear how this will impact the treaty's effectiveness. The features that make it politically feasible and more robust as an instrument of AI regulation are also those that could weaken its practical impact. These include, for example, creating a blind spot for potentially harmful AI applications and leading to a situation of uneven implementation across jurisdictions.

Potential for Attracting Signatories

The Framework Convention's effectiveness also depends on its ability to attract and retain signatories. Greater participation will give it more regulatory weight and will contribute to less fragmentation in global AI regulation. This treaty may be particularly well-positioned to attract countries without existing AI regulation experience or expertise, as it offers a useful entry point into AI governance, legitimacy, and guidance without heavy administrative costs (Interview 2). Further, its flexible design and focus on the shared principles of human rights, democracy, and rule of law make it attractive for states to draw on when shaping their own AI regulations without being bound to imitate approaches taken by other states that may be domestically unfeasible (Interview 6).

That the Framework Convention can count important AI actors among its signatories is a positive sign. Yet, without a critical mass of signatories and a global uptake, its potential for influencing regulatory behaviour and achieving impact will remain limited. The CAI has prioritised outreach to Latin American and African countries, which has led to Uruguay signing the treaty in September 2025 and states such as Chile, Ecuador, Brazil, Ghana, South Korea, and Cameroon expressing an interest in becoming observers (Interviews 2 and 6). The CoE Secretariat has also taken proactive steps to promote the Convention's uptake, with the Director for Human Rights, Hanne Juncher, discussing it with the African Union (AU) in November 2024 during the Cairo-based OECD-AU AI Dialogue (Bureau Report 2024).

A critical test for the treaty is whether major global AI actors beyond Europe also ratify it. Despite being a signatory and having played a strong role in the negotiation of the treaty, it is highly unlikely that the US will ratify the Convention under the Trump administration (Interview 1). This may make it less likely for the treaty to gain traction internationally, which would ultimately reduce its effectiveness, robustness, and democratic attributes. Yet, if even some key states ratify, others may follow out of a desire to have a seat at the table and to impact both the initial design and the decisions of the Conference of the Parties follow-up mechanism, as discussed below (Interview 6).

Implementation

The treaty's effectiveness will also depend on domestic implementation of actions that comply with its obligations. States that ratify the treaty will be obliged to transpose its provisions into law, but because the Convention is framed in broad principles, there is significant scope for variation (Schneider 2024). This will inevitably lead to inconsistencies in how the treaty will be implemented across different jurisdictions. To address this, the Framework Convention is accompanied by HUDERIA to help translate the legal obligations of the treaty into a structured, risk- and impact-assessment methodology. It asks actors to evaluate how AI systems affect fundamental rights and democratic institutions, and to document and mitigate risks (Council of Europe 2024c). HUDERIA itself is non-binding, but it provides states with a practical model for implementation, increasing the likelihood of a more consistent implementation and reducing the risk of fragmented assessment frameworks — both of which can contribute to improving the treaty's effectiveness.

We may see smaller states lacking the resources to build effective oversight bodies or larger states opting for more minimalist approaches.

HUDERIA has already been piloted with CoE members and observers. The CoE also ran the "HUDERIA Academy," a training session in June 2025 that involved 30 countries and 100 participants, including small- and medium-sized enterprises (Interview 5). In addition to building capacity amongst public administrations for assessing AI-related risks, it also provided an introduction to understanding human rights for private sector participants and created a forum for building a community of practice around aligning AI with human rights, democracy, and rule of law (Interview 2). At present, it looks likely that these kinds of capacity-building activities will continue to be part of the mandate of the new committee that will take over responsibility from the CAI starting in 2026 (Interview 6).

The treaty also requires that parties establish supervisory bodies and grievance mechanisms, and, by embedding such procedural rights, it attempts to make its principles actionable. The effectiveness of these provisions will, however, depend on national capacity to actually implement them. Instruments like HUDERIA can help, but the degree of uptake on these provisions will likely vary. We may see, for example, smaller states lacking the resources to build effective oversight bodies or larger states opting for more minimalist approaches.

Follow-up and Compliance

Another key pillar of effectiveness is the design of, and participation in, the compliance and follow-up mechanisms. Here, there are two key considerations. First, once the treaty enters into force, a Conference of the Parties (CoP) will be created and tasked with overseeing implementation, exchanging information, reviewing interpretative questions, and potentially adopting protocols or amendments (Council of Europe 2024a). This is important, as efforts in other areas of global digital regulation show that effectiveness is not only a function of bindingness, but also of promoting information-sharing and building synergies between actors in this space (Marconi and Greco 2025). In theory, the CoP could become a highly useful forum where state and non-state actors collectively respond to new AI risks, issue guidance, build capacity for states' AI governance approaches, and perhaps even strengthen obligations. How helpful the CoP will be in practice will depend on its design. Since the treaty text leaves its rules of procedure to be set by states after entry into force, early ratifiers will prove decisive in shaping the scope of its role.

Effectiveness will also depend on how the CAI gets extended or transformed after its mandate expires at the end of 2025.

Second, effectiveness will also depend on how the CAI gets extended or transformed after its mandate expires at the end of 2025. Discussions are underway about building on the work of the CAHAI and the CAI by creating a standing committee on AI and human rights within the CoE (Interview 3). This committee would serve as a continuation of the CAI; it could monitor developments, issue recommendations to the CoP, and relaunch negotiations if adaptations to the Framework Convention are needed or desirable (Interviews 3 and 6). The committee would also ensure continuity of expertise and provide a platform to promote the treaty and attract new signatures. Those observers already admitted to the CAI would retain permanent status, though how exactly civil society and private sector participation will look remains uncertain. As these procedures were not laid out explicitly in the treaty text, much is left to the discretion of states (Interview 2).

Democracy

The democratic quality of the Framework Convention can be assessed along three dimensions: (1) participation and inclusivity, (2) transparency and accountability, and (3) its fidelity to democratic values more broadly. Within each of these dimensions, the Framework Convention makes notable contributions, but also reveals some weaknesses.

Participation and Inclusivity

The negotiation process included all 46 CoE members, as well as observers from other states, regional and international organisations, AI experts, and representatives from CSOs and the private sector (Schneider 2024; Interviews). The Framework Convention is universally open to accession, with any state able to become a full party to the treaty with equal rights in

the CoP (Council of Europe 2024a). In theory, this creates the possibility for the Framework Convention to emerge as a global forum for AI regulation.

In practice, however, the negotiation process fell short in terms of participation and inclusiveness. Key actors in AI development, such as China, India, and Russia, were not present, and representation from the Global South was limited, creating a legitimacy gap. Further, certain CoE observer states, particularly the US with its technical expertise and political leverage, wielded disproportionate influence, despite lacking voting rights within the CoE (Bertuzzi 2024). The role of non-state actors was equally mixed: 68 representatives from civil society, academia, and industry were formally involved in the negotiation process and were able to participate in the CAI's plenary sessions. However, at the suggestion of governments, they were excluded from the treaty drafting group, undermining transparency and going against the CAI's terms of reference that mandated the meaningful participation of civil society.

Transparency and Accountability

Democratic legitimacy depends on transparency and accountability in both the negotiations and the follow-up of the Framework Convention. Here, the record is also mixed. The exclusion of non-state participants from drafting sessions, combined with the confidentiality of negotiations, raised concerns about transparency (Bertuzzi 2024; Interview 7). Further, the Framework Convention's procedural rights, remedies, and safeguards aim to foster accountability at the domestic level; states must ensure that individuals interacting with AI systems are notified, can access information about the system, and can challenge AI-based decisions before competent authorities (Council of Europe 2024a). These requirements strengthen the rule of law and enable citizens to hold governments accountable. While the Framework Convention does not create new rights, it provides courts with clear benchmarks for national judges and, indirectly, the ECHR to interpret its principles when applying existing human rights law (Interview 3).

The CoP will be crucial for ensuring democratic accountability at the international level.

Looking ahead, the CoP will be crucial for ensuring democratic accountability at the international level. If designed inclusively rather than symbolically, it could serve as a venue for civil society to scrutinise government compliance (Interview 3). The CoE has already started working on capacity-building efforts through HUDERIA trainings, illustrating its intent to create an international forum where information on AI risks, regulatory practices, and implementation challenges can be exchanged. However, access to information around the Framework Convention will depend on how ambitiously states design the CoP's mandate and how detailed initial disclosures are on compliance. Early ratifiers will likely set precedents for how far disclosures go, and this will certainly influence the direction of the CoP going forward (Interview 1). On the one hand, extensive disclosure of compliance could be one of the treaty's most tangible contributions and could lead to higher effectiveness. On the other hand, setting a high bar for disclosure (and thus scrutiny) could also scare away potential signatories or ratifiers, thus jeopardising the treaty's effectiveness, robustness, and democratic quality.

Normative Alignment with Democratic Values

The Framework Convention enshrines democratic governance values directly in its text. It reaffirms commitments to human rights, equality, and non-discrimination and makes these principles binding in the context of AI (Council of Europe 2024a). Further, by requiring states to legislate oversight, ensure transparency, and establish grievance mechanisms, the treaty compels governments to embed democratic processes into AI regulation. The Framework Convention also provides a dedicated platform of exchange on AI for states committed to human rights, democracy, and the rule of law, setting it apart from broader forums like the G20 or the OECD. As witnessed in the past with the Budapest Convention on Cybercrime, the Framework Convention has the potential to draw in a wide set of countries, build capacity, and set global standards for aligning AI regulation with human rights and democratic values (Interview 1).

The Role of the EU

The EU, represented by the European Commission, and its member states played a decisive role in shaping the Framework Convention, both in ideological and practical terms. But the road to get there, and particularly the EU's role within the negotiations, was not without tension. The Commission negotiated on behalf of all EU member states and tightly managed coordination, evoking the duty of sincere cooperation to ensure unity among EU member states (Interview 3). At times, this strategy created friction with other delegations and the CAI, as the EU delayed negotiations until the parallel process of legislating the AI Act had proceeded far enough for the Commission to finally secure a negotiation mandate (Interviews 1, 4, and 6).

Once a mandate was secured, the Commission was able to have a significant impact on the treaty. Most importantly, its influence took the form of internationalising rights-based principles of AI regulation and successfully inserting its own regulatory DNA into the treaty by linking it strongly to the concepts of the AI Act (Interview 4). These include, for example, the principle of differentiated obligations proportional to risks, a human-centric framing that emphasises fundamental rights, human dignity, and democratic values, the notion of transparency requirements to ensure that individuals are informed when they interact with AI systems and have access to sufficient information to challenge AI-driven decisions, and finally, provisions for responsible innovation through regulatory sandboxes. The EU also reinforced the relevance and potential effectiveness of the Framework Convention by using it as a complement to the AI Act, which serves as a ready-made compliance mechanism for the 27 EU members.

The EU also reinforced the relevance and potential effectiveness of the Framework Convention by using it as a complement to the AI Act.

The EU's central negotiating position was to maximise the treaty's international relevance, even if doing so would render the final instrument less stringent (Interview 4). This choice reflected a deliberate pursuit of robustness. By prioritising flexibility with the aim to attract the widest possible set of signatories, the EU helped strengthen the treaty's legitimacy and reach beyond its own member states. While this approach did indeed attract additional signatories (chiefly, the US), this robustness came at the expense of effectiveness, as key obligations were softened or left open to national discretion. With regard to democracy, this approach diluted some of the stronger protections and oversight mechanisms that European civil society actors had called for.

The EU aimed to increase the treaty's effectiveness by including the regulation of both public and private sector use of AI by default — a move in line with the AI Act. The European Commission strongly resisted US demands to exclude companies from the treaty's obligations (Murphy and Jacobson 2024), arguing that AI technologies posed risks to human rights, democracy, and the rule of law, regardless of whether the public or private sector deployed or used it, and that any regulation exclusively aimed at the public sector would be too limited (Interview 4).

Even though the final text adopted the opt-in mechanism favoured by the US, the UK, Japan, and Canada, the EU's resistance to the exclusion of the private sector ensured that private sector actors working on behalf of public bodies are covered by the treaty.[7] What's more, risks originating from private sector uses of AI are explicitly acknowledged in the treaty's provisions and states have to explain how they intend to address these risks upon ratification. Despite not achieving full default coverage of companies, the European Commission's critical voice prevented the treaty from excluding private sector uses of AI entirely.

The European Commission's critical voice prevented the treaty from excluding private sector uses of AI entirely.

The EU was more successful in shaping how the treaty deals with national security. Here, the Commission firmly held that the Framework Convention must follow the AI Act's approach, which excludes national defence and security from its remit. This approach was politically necessary given that EU member states have sovereignty over security matters. While this carve-out contributed to the robustness of the treaty by making it more politically acceptable to a broader group of signatories, it reduced effectiveness by exempting some of the most rights-sensitive applications of AI under the guise of national security. From a democracy perspective, this carve-out also limits the ability of treaty participants, particularly non-state observers, to scrutinise precisely those uses of AI that can acutely affect fundamental rights.

The EU's role in making the treaty more democratic was similarly mixed. Although the Commission initially favoured an open drafting process that would include civil society actors, some EU member states sided with the UK and US in pushing for confidentiality and the Commission relented (Interviews 3 and 4). Nevertheless, during the plenary sessions, EU representatives did engage with CSOs' comments, explained the Commission's position, and, where possible, accommodated CSO proposals. The EU, for example, openly supported calls to integrate environmental protection and sustainability assessments into the explanatory report (Interviews 3 and 4).

Going forward, the EU has an opportunity to provide leadership in the treaty's follow-up phase. The CoP, as well as the follow-up committee to the CAI, offer the EU platforms for exporting its standards and interests, as well as venues for building coalitions to enhance the Framework Convention's reach and relevance. In the CoP, EU members will likely form the largest bloc of early ratifiers, giving the European Commission and its member states significant influence in shaping its rules and design. This bloc could set precedents regarding ensuring greater transparency and non-state stakeholder participation, thus making the mechanism more democratic. The EU could also play a pivotal role in shaping the Framework Convention's trajectory and improving its effectiveness by doubling down on promoting the AI Act internationally and using the CoP as a platform to

7 Perhaps made easier at the time due to the fact that this was also covered the Biden Administration's Executive Order on AI and that the U.S. executive branch has the legal authority from the U.S. Congress to set the terms of how private entities working on behalf of the government (e.g. government contractors) develop and use AI.

align global practice with European standards (Interview 1). Alternatively, the EU could stick with a more flexible, loose approach and avoid setting high precedents with regard to implementation and reporting back. This may make it more attractive to additional signatories wary of high compliance needs and scrutiny, and thus shore up its robustness.

With the eventual ratification of the treaty, the EU will have constructed for itself a layered system of AI governance, with the Framework Convention at the level of principles, the AI Act at the level of detailed regulation, and HUDERIA as an implementation methodology. This tiered structure serves as a case study in how a regional actor can drive global governance discussions. It also illustrates a model that other states or regions may find worth emulating.

The challenge will be whether the EU can sustain its leadership in a shifting geopolitical landscape, generally and — specifically, with regard to the Framework Convention and global AI governance — balance the robustness and democratic legitimacy of broad participation with the effectiveness of strong, enforceable protections.

Conclusion

Negotiated under time pressure and in the shadow of rapidly shifting global narratives around AI, the Framework Convention on AI very much reflects the trade-offs and geopolitical realities of the present. Nevertheless, it is an impressive achievement. The treaty shows that multilateral approaches can work even as we see an erosion of the idea of a global, open internet and the emergence of a more fragmented space, emphasising digital sovereignty, economic competitiveness, and national security. Coming to an agreement, even among a group of largely like-minded states, on a binding instrument that embeds human rights, democracy, and rule of law principles into global AI governance is an important step forward.

The negotiations succeeded in producing a legally binding instrument that regulates AI across its lifecycle.

As with its earlier treaties on cybercrime and data protection, the CoE brought institutional expertise, convening power, treaty management infrastructure, and a rights-based framework to the table. In the end, it paid off: the negotiations succeeded in producing a legally binding instrument that regulates AI across its lifecycle. In doing so, the CoE created a global AI governance anchor that harmonises with existing global, non-binding initiatives such as the OECD AI Principles and the UNESCO Recommendation on the Ethics of Artificial Intelligence.

The compromises that made the agreement possible also limit its potential for effectiveness. The Convention deliberately adopted a principles-based and technology-neutral design to ensure robustness across time and to attract more signatories; this helped strengthen its resilience, but it came at the cost of limiting the treaty's effectiveness. The opt-in approach to private-sector obligations and the carve-out for national security leave notable gaps precisely where the most rights-sensitive applications of AI lie. Yet, as global narratives continue to move away from cooperative calls for global regulation and toward national security and competitiveness, the treaty's effectiveness and robustness will likely be impacted, risking its ability to gain traction as a truly global instrument.

From a democratic standpoint, while negotiations around the treaty show a mixed record that highlights the tension between greater participation and expediency, the end result holds promise. On the one hand, the negotiations fell short with an under-representation of voices from the Global South, the exclusion of non-state actors from the drafting group, and the requirement of confidentiality of draft texts. On the other hand, the CAI provided draft texts to non-state observers for comment, and it published an explanatory report providing useful additional context on the negotiations and the interpretation of the final text. In addition, the CoP mechanism is mandated to allow multi-stakeholder engagement and this should, if done meaningfully, strengthen accountability and its democratic credentials.

Importantly, the Framework Convention is now open to accession, and it gives all signatories who ratify it an equal voice going forward. While oversight over compliance with the treaty is largely left to national jurisdictions and their domestic supervisory bodies, the treaty does

embed certain judicial avenues via the European Court of Human Rights' interpretive role. The CoP will also provide a forum for follow-up on whether and how states are meeting their obligations. How the CoP is designed and implemented, and what the role of a post-CAI committee looks like, will be a test of the treaty's democratic credentials going forward.

The EU used the treaty as an instrument to channel its AI Act principles to the international level, successfully embedding risk-based, human-centric obligations into the Framework Convention. This contributed to the treaty's robustness by offering a coherent template others could adapt and to its effectiveness, as EU member states can simultaneously fulfil treaty obligations via their AI Act compliance. While the EU's compromises on the opt-in model for private actors and the national security carve-out led to what will surely be a less effective treaty, they contributed to a more globally palatable and, as a result, more robust treaty. By signing the treaty (and likely ratifying it), the EU has built a layered and harmonised model of AI regulation — made up of the Framework Convention for top-level principles, the EU AI Act for direct regulation, and the HUDERIA methodology for implementation — that serves as an interesting example for others to potentially replicate.

The EU used the treaty as an instrument to channel its AI Act principles to the international level.

The Framework Convention and the negotiation process offer some important lessons for other multilateral efforts to regulate AI systems. It demonstrated how difficult it is to achieve consensus on binding rules even among a collection of largely like-minded states; compromises were needed that limited ambition. In addition, the experience of the CoE's Convention on Cybercrime shows that strong regional instruments can trigger parallel, weaker initiatives at the UN level, as states opposing stringent standards seek to dilute their impact (Interview 3). This suggests that the global efforts around, for example, the United Nations Global Digital Compact, where a wider range of political systems and values exist, are unlikely to yield meaningful AI regulation in the near term.

In addition, the Framework Convention process highlights both the advantages and limitations of what, in principle, should be the smoother path of working with like-minded states. While this approach led to the adoption of a legally binding treaty, the absence of stakeholders from, for example, important AI actors such as China or the Global South, raises questions of legitimacy that UN processes will not face. Other, more global efforts to regulate AI will face a similar trade-off between effectiveness, robustness, and democracy.

It is unlikely that the Framework Convention will become the dominant global framework for AI regulation. Instead, it will likely coexist with a patchwork of regional regulations, soft-law principles, and national or sub-national legislation. Yet, its deliberate design may also prove to be a strong point with regard to, for example, shaping judicial interpretation in Europe or offering a model for other countries for their own domestic AI regulation. Similar to the Budapest Convention on Cybercrime that eventually set global standards, the AI Convention could, over time, achieve a similar reach if it gains ratifications and proves useful in practice.

Whether the Framework Convention becomes a useful tool for global AI regulation hinges on three key factors. First, while it requires only five ratifications to enter into force, its authority, as well as its effectiveness, robustness, and democratic credentials, will depend on whether major players outside Europe choose to ratify it. Without them, the treaty risks being a purely European instrument with global aspirations but little legitimacy. Right now, such an outcome seems likelier than it did when negotiations started in 2022. The treaty's long-term robustness and effectiveness will depend less on its legal design and more on whether current geopolitical developments are conducive to sustained engagement. To maximise the Framework Convention's potential as a viable instrument, the EU should quickly ratify it.

The Framework Convention's authority will depend on whether major players outside Europe choose to ratify it.

Second, the treaty's principles need translation into national laws and practices, a process which can be difficult given the complexity of AI systems and the speed with which they are developing. Here, tools like HUDERIA can help by providing a structured methodology for assessing AI's impact on human rights, democracy, and the rule of law, and for designing regulatory systems. To support this, the CoE, the CAI, and the new post-CAI committee should continue with their efforts to promote HUDERIA through trainings and pilot projects. These efforts can also serve as an entry point for potential new signatories. Eventual parties to the treaty, including the EU, should support these efforts through offering financial and technical resources.

Finally, the design of the CoP and the role of a post-CAI committee will be crucial for follow-up and information sharing. If early ratifiers (likely led by the EU) design strong procedures, ensure transparency, and allow for meaningful non-state and expert input in the context the CoP, and a post-CAI committee proves to be an active forum for discussion around new AI developments and other emerging technologies, the Framework Convention could evolve into a dynamic instrument that is more effective, robust, and democratic.

List of Interviews

Number	Date	Interviewee	Location
1	07/03/2025	Government participant in the CAI negotiations	Online
2	07/14/2025	Member of the CAI Bureau	Online
3	07/15/2025	CSO representative	Online
4	07/16/2025	European Commission Representative	Online
5	7/22/2025	AI Expert	Online
6	7/24/2025	Member of the CAI Bureau	Online
7	8/22/2025	Private sector representative	Online

References

Access Now. 2023. *Civil Society Statement on the Council of Europe Convention on AI*, January 28. https://www.accessnow.org/wp-content/uploads/2023/01/CSO-COE-CAI-Statement-28012023.pdf.

AlgorithmWatch. 2024. "The Council of Europe's Convention on AI: No Free Ride for Tech Companies and Security Authorities!" *AlgorithmWatch* CH, January 29. https://algorithmwatch.ch/en/council-of-europes-ai-convention-no-free-ride/.

Bazoobandi, Sara, Sangeeta Mahapatra, Tetiana Schipper, and Iris Wieczorek. 2025. "From Global Governance to Nationalism: The Future of AI." *GIGA Focus Global*, no. 2. Hamburg: German Institute for Global and Area Studies (GIGA). https://doi.org/10.57671/gfgl-25022.

Bertuzzi, Luca. 2023. "EU's AI ambitions at risk as US pushes to water down international treaty." *Euractiv*, June 6. https://www.euractiv.com/news/eus-ai-ambitions-at-risk-as-us-push-to-water-down-international-treaty/.

—. 2024. "EU Commission's Last-Minute Attempt to Keep Private Companies in World's First AI Treaty." *Euractiv*, April 18. https://www.euractiv.com/news/eu-commissions-last-minute-attempt-to-keep-private-companies-in-worlds-first-ai-treaty/.

Bremmer, Ian, and Mustafa Suleyman. 2023. "The AI power paradox: Can states learn to govern artificial intelligence-before it's too late?." *Foreign Affairs* 102: 26.

Buhmann, Alexander, and Christian Fieseler. 2021. "Towards a Deliberative Framework for Responsible Innovation in Artificial Intelligence." *Technology in Society* 64: 101475.

Caunes, Parrón. 2023. "Civil Society Urges Coverage of All AI Systems, Opposes Exemptions." LinkedIn. https://www.linkedin.com/pulse/civil-society-urges-coverage-all-ai-systems-opposes-parr%C3%B3n-phd-vcfef/.

Choi, Ha Eun, Hylke Dijkstra, Andrea Liese, Thomas Sommerer, and Clara Weinhardt. 2024. "EU Support for Robust, Effective, and Democratic Global Governance: A Conceptual Framework." *ENSURED Research Report*, no. 4 (July): 1–44. https://www.ensuredeurope.eu.

Cole, Mark. 2024. "AI Regulation and Governance on a Global Scale: An Overview of International, Regional and National Instruments." *AIRe - Journal of AI Law and Regulation* 1 (1): 126–142. https://doi.org/10.21552/aire/2024/1/16.

Collingridge, David. 1980. *The Social Control of Technology*. New York: St. Martin.

Council of Europe. 2019a. 1353rd *Meeting of the Ministers' Deputies: Minutes (Item 1.5 — Council of Europe Work on Artificial Intelligence, Terms of Reference for the Ad hoc Committee on Artificial Intelligence [CAHAI])*, September 11. Strasbourg: Council of Europe. https://search.coe.int/cm?i=0900001680972f20.

—. 2019b. *Council of Europe Work on Artificial Intelligence*. Information Documents SG/Inf(2019)21, July 02. Strasbourg: Council of Europe.

—. 2020a. *Feasibility Study*. https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-/1680a0c6da.

—. 2020b. *Towards Regulation of AI Systems*. Strasbourg: Council of Europe. https://rm.coe.int/prems-107320-gbr-2018-compli-cahai-couv-texte-a4-bat-web/1680a0c17a.

—. 2021. *Analysis of the Multi-Stakeholder Consultation* (CAHAI(2021)07). Strasbourg: Council of Europe. https://rm.coe.int/cahai-2021-07-analysis-msc-23-06-21-2749-8656-4611-v-1/1680a2f228.

—. 2021b. *Possible Elements of a Legal Framework on Artificial Intelligence, Based on the Council of Europe's Standards on Human Rights, Democracy and the Rule of Law*. Report CM(2020)90-final, September 24. Strasbourg: Council of Europe.

—. 2022. *Council of Europe's Digital Partnership*. Strasbourg: Council of Europe. https://rm.coe.int/10-06-2022-digital-partnership-general-doc-updated/1680a6e634.

—. 2024a. *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*, CETS, no. 225. Strasbourg: Council of Europe. https://rm.coe.int/1680afae3c.

—. 2024b. *Explanatory Report to the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*, CETS, no. 225. Strasbourg: Council of Europe. https://rm.coe.int/1680afae67.

—. 2024c. *Methodology for the Risk and Impact Assessment of Artificial Intelligence Systems from the Point of View of Human Rights, Democracy and the Rule of Law (HUDERIA Methodology)*. Strasbourg: Council of Europe.

Council of the European Union. 2022. *Council Decision (EU) 2022/2349 of 21 November 2022 authorising the opening of negotiations on behalf of the European Union for a Council of Europe convention on artificial intelligence, human rights, democracy and the rule of law*. Official Journal of the European Union L 311/138 (December 02).

Hofmann, Staphanie and Patryk Pawlak. 2024. "Harnessing Europe's Narrative Power to Shape the Digital Future." *Carnegie Endowment for International Peace*, October 10. https://carnegieendowment.org/research/2024/10/harnessing-europes-narrative-power-to-shape-the-digital-future?lang=en.

International Chamber of Commerce. 2023. *ICC comments on the Council of Europe draft Convention on Artificial Intelligence*. https://icc.se/wp-content/uploads/2023/11/ICC_CoEAI_ConsolidatedConvention_Comments_Final-1.pdf.

Juncher, Hanne. 2025. "Statement at the CPDP 2025 Panel: 'Assessing Fundamental Rights Risks of AI: Opportunities and Challenges Ahead.'" YouTube video, posted by CPDP Conferences, June 4.

Leslie, David, Christopher Burr, Mhairi Aitken, Josh Cowls, Michael Katell, and Morgen Briggs. 2021. *Artificial Intelligence, Human Rights, Democracy, and the Rule of Law: A Primer*. Strasbourg: Council of Europe.

Marconi, Federica and Ettore Greco. 2025. "Regulating Cyberspace: UN Consensus-Building in a Fragmented Digital World." *ENSURED Research Report*, no. 10 (May): 1–34. https://www.ensuredeurope.eu.

Murphy, Claire, and Julia Jacobson. 2024. "Heavyweight Fight, Did the US or EU KO the AI Treaty?" *Privacy World (blog)*, April 17. https://www.privacyworld.blog/2024/04/heavyweight-fight-did-the-us-or-eu-ko-the-ai-treaty.

OECD. 2024. *AI Principles*. https://www.oecd.org/en/topics/ai-principles.html.

—. 2025. *The OECD.AI Policy Navigator*. https://oecd.ai/en/dashboards/policy-initiatives.

Volpicelli, Gian. 2024. "Council of Europe Makes a Mockery of International AI Rights Treaty." *Politico Europe*, March 11. https://www.politico.eu/article/council-europe-make-mockery-international-ai-rights-treaty.

Powell, Rosamund. 2024. "The EU AI Act: National Security Implications." *CETaS Explainer* (August): 1–12. https://cetas.turing.ac.uk/publications/eu-ai-act-national-security-implications.

Roberts, Huw, Emmie Hine, Mariarosario Taddeo, and Luciano Floridi. 2024. "Global AI Governance: Barriers and Pathways Forward." *International Affairs* 100 (3): 1275–1286.

Roche, Cathy, Dave Lewis, and P.J. Wall. 2021. "Artificial Intelligence Ethics: An Inclusive Global Discourse?" In *Proceedings of the 1st Virtual Conference on Implications of Information and Digital Technologies for Development*, 643–658.

Schneider, Thomas. 2024. "Regulating AI and Digital Technologies – What the New Council of Europe Convention Can Contribute." In *Artificial Intelligence and the Challenge for Global Governance: Nine Essays on Achieving Responsible AI*, edited by Alex Krasodomski. London: Chatham House. https://www.chathamhouse.org/2024/06/artificial-intelligence-and-challenge-global-governance/03-regulating-ai-and-digital.

Taggart, Jack, and Sebastian Haug. 2024. “De Jure and de Facto Inclusivity in Global Governance.” *Review of International Studies*: 1–25. https://doi.org/10.1017/S0260210524000627.

United Nations. 2025. *Global Digital Compact*. https://www.un.org/global-digital-compact/en.

UNESCO. 2023. *Recommendation on the Ethics of Artificial Intelligence*. https://www.oecd.org/en/topics/ai-principles.html.

Wallach, Wendell and Gary Marchant. 2019. “Toward the Agile and Comprehensive International Governance of AI and Robotics.” *Proceedings of the IEEE* 107 (3): 505–508.

White House. 2023. *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*. Washington, DC.

Zhang, Hongzhou, Shaleen Khanal, and Araz Taeihagh. 2025. “Public-Private Powerplays in Generative AI Era: Balancing Big Tech Regulation Amidst Global AI Race.” *Digital Government: Research and Practice* 6 (2): 1–11. https://doi.org/10.1145/3664824.

About ENSURED

In an era marked by global challenges, international cooperation is more essential than ever. Yet multilateral initiatives too often end in gridlock, as dominant states seek to bend the global order to their own interests. Enter ENSURED, a Horizon Europe-funded research consortium studying how the EU and its member states can better defend multilateralism and make global governance more robust, effective, and democratic.

ENSURED focuses on key policy domains that by their very nature pose complex transnational challenges. Our research assesses the state of play in these different areas and investigates the EU's strengths and weaknesses as an actor working to defend and transform multilateralism. Embracing the ethos of multilateral cooperation, the ENSURED consortium comprises universities, think tanks, and civil society groups from across Europe, Brazil, India, South Africa, China, and the United States. We aim to equip policymakers in the EU with evidence-based insights, actionable recommendations, and practical tools to promote better global governance for a world in transition.

ENSURED publications are available via the project website: https://www.ensuredeurope.eu/

Funded by the European Union

The ENSURED project is funded by the European Union's Horizon Europe research and innovation programme under the Call HORIZON-CL2-2022-DEMOCRACY-01 – Grant agreement n° 101092077. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (granting authority). Neither the European Union nor the granting authority can be held responsible for them.

Edited by: Oliver Jung
Editorial coordination: Global Public Policy Institute (GPPi)

Reinhardtstr. 7
10117 Berlin
Germany
ensured@gppi.net