CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Council of Europe Framework Convention on Artificial Intelligence: Context, Regulatory Approach and Scope of Obligations

Vladislava Stoyanova[iD]

Faculty of Law, Lund University, Lund, Sweden
Email: vladislava.stoaynova@jur.lu.se

Abstract

The Council of Europe has very recently adopted the Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law. This article provides an initial analysis of the CoE AI Convention. It emphasises the necessity of understanding the CoE AI Convention within the context of its adoption as an international treaty negotiated within the Council of Europe. This context has affected its scope in terms of how the treaty includes the regulation of the usage of AI systems by both public authorities and private actors. The detailed review of the available negotiation documents reveals that the concrete level of protection offered by the Convention has been lowered. This includes the risk-based approach, which shapes the obligations undertaken by States under the treaty. This approach is explained and contrasted with the approach under the EU AI Act. The argument that emerges is that the absence of categorisation of risk levels in the treaty is related to its higher level of abstraction, which does not necessarily imply less robust obligations. The content of these obligations is also clarified in light of the requirement imposed by the treaty of consistency with human rights law. An argument is advanced that the principles formulated in the treaty – human dignity and individual autonomy, transparency and oversight, accountability, non-discrimination, data protection, reliability, risk-management – can offer interpretative guidance for the development of human rights standards.

Keywords: AI; Council of Europe; human rights

I. Introduction

The risks that Artificial Intelligence (AI) systems pose have prompted States to undertake regulatory measures. The EU adopted its Regulation on Artificial Intelligence (hereinafter AI Act),[1] which has been heralded as the first comprehensive framework intended to regulate AI.[2] The AI Act has already entered into force and while the implications from many of its provisions are yet to be better understood,[3] the AI Act has already been an

---

[1] Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), OJ L 2024/1689.

[2] See, e.g., F Palmiotto, "The AI Act Roller Coaster: The Evolution of Fundamental Rights Protection in the Legislative Process and the Future of the Regulation" (2025) 16(2) European Journal of Risk Regulation 770, where further references are also provided.

[3] See Art. 113 for the varying timeframes of applicability of the different provisions has the AI Act.

object of useful commentaries and analysis.[4] Similarly to the EU, the Council of Europe (hereinafter CoE) has also very recently adopted the Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (hereinafter the CoE AI Convention).[5] At the time of writing the treaty has not entered into force.[6] It has been signed by eleven CoE Member States and by the EU, the United States of America, Canada, Japan, Uruguay and Israel. No State has ratified it yet.[7]

The CoE AI Convention defines AI system as "machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations or decisions that may influence physical or virtual environments." This definition aligns with the one enshrined in Article 3(1) of the AI Act. It has been considered relatively wide and meant to keep up with future technological developments.[8] With this definition in mind, this article provides an initial analysis of the CoE AI Convention. Four steps are followed. The first emphasises the necessity of understanding the CoE AI Convention within the context of its adoption as an international treaty negotiated within the CoE (Section II). This context has affected the personal scope of the treaty in terms of how it includes the regulation of the usage of AI systems by both public authorities and private actors. A detailed review of the available negotiation documents reveals that the concrete level of protection provided by the CoE AI Convention has been reduced, not least through the drafters' final decision to regulate private actors only indirectly. Section III explains this personal scope of the Convention, including by describing the drafting negotiations concerning the regulation of public authorities and private actors. This regulation follows a risk-based approach, which shapes the obligations undertaken by States under the treaty. Section IV explains the risk-based approach and contrasts it with the approach under the EU AI Act. The argument put forward is that the treaty's omission of explicit risk categories reflects its higher level of abstraction, which does not, however, entail weaker obligations. Section V explains the principles articulated in the Convention and intended to shape the obligations undertaken by States. It is argued that these principles, namely human dignity and individual autonomy, transparency and oversight, accountability, non-discrimination, data protection, reliability and risk management, can serve as interpretative tools for shaping the development of human rights standards.

A note on the methodology is due. As an international treaty, the CoE AI Convention is an object of treaty interpretation in accordance with the rules of interpretation articulated in Articles 31 and 32 of the Vienna Convention on the Law of Treaties (Vienna Convention).[9] Pursuant to Article 31(1) of the Vienna Convention, the treaty is to be interpreted "in accordance with the ordinary meaning to be given to the terms of the

---

[4] C N Pehlivan, N Forgó and P Valcke (eds), *The EU Artificial Intelligence (AI) Act: A Commentary* (Kluwer 2024); J Schuett, "Risk Management in the Artificial Intelligence Act" (2024) 15(2) European Journal of Risk Regulation 367; O Mir, "The AI Act from the Perspective of Administrative Law: Much Ado About Nothing?" (2024) 16(1) European Journal of Risk Regulation 63; I Carnat, "Addressing the Risks of Generative AI for the Judiciary: The Accountability Framework(s) under the EU AI Act" (2024) 55 Computer Law & Security Review 106067; E Leinarte, "The Classification of High-Risk AI Systems Under the EU Artificial Intelligence Act" (2024) 1 Journal of AI Law and Regulation 262.

[5] CETS No. 225 (Vilnius 05/09/2024).

[6] See Full list – Treaty Office.

[7] To enter into force five ratifications including at least three by Member States of the CoE, are required.

[8] K Stuurman and E Lachaud, "Regulating AI. A Label to Complete the Proposed Act on Artificial Intelligence" (2022) 44 Computer Law and Security Review 1. It is necessary to also await the Commission's clarification of the concept of AI system in the guidelines on the practical implementation of the AI Act to be drawn up (see Art. 96(1)(f) AI Act). See also Miguel Ángel Presno Linera and Anne Meuwese, "Regulating AI from Europe: a Joint Analysis of the AI Act and the Framework Convention on AI" (2025) The Theory and Practice of Legislation 1, 10.

[9] Vienna Convention on the Law of Treaties of 23 May 1969, entered into force on 27 January 1980. United Nations, Treaty Series, vol. 1155, p. 331. I approach treaty interpretation as an *interactive* process structured on the

treaty in their context and in the light of its object and purpose." Article 1(1) of the CoE AI Convention defines the object and purpose of the treaty as ensuring "that activities within the lifecycle of artificial intelligence systems are fully consistent with human rights, democracy and the rule of law." Human rights law therefore forms an integral part for interpreting the treaty, which justifies the references to human rights in the forthcoming analysis. No attempt is undertaken to offer a comprehensive analysis as to how the provisions of the treaty could be interpreted in light of human rights law. In fact, human rights law standards are yet to develop with some higher level of precision in this area.[10] The principles expressed CoE AI Convention can help in the development of human rights law.[11] In this sense, a mutual interaction between the treaty and human rights standards can be expected.[12]

According to Article 31(3)(c) of the Vienna Convention, "any relevant rules of international law applicable in the relations between the parties" can be also taken into account for interpreting the treaty. EU law and in particular the EU AI Act can be considered as a source of relevant rules. An integrated interpretative approach is therefore followed whereby not only human rights law is relevant, but more generally CoE and EU law need to be understood in light of each other.[13] There is indeed a plurality of sources that operate at European level,[14] and they need to be interpreted in an integrated way.[15] While no attempt for a comprehensive comparison is offered in this article, the EU AI Act does provide important points of reference for better understanding the CoE AI Convention. Section IV offers a more detailed comparison with the EU AI Act regarding the conditions for the classification of systems as high-risk under the Act. The reason is that this very classification can be considered as a key point of contrast at a more general conceptual level in terms of how AI systems have been regulated at European level (i.e., by the CoE and by the EU). This justifies a more in-depth comparative analysis, as offered by Section IV.

As it is the case with other CoE treaties,[16] the Explanatory Report published together with the treaty is an important source for better understanding how these obligations

---

basis of the text, context, object and purpose of a treaty and guided by good faith. See Campbell McLachlan, *Principle of Systemic Integration in International Law* (Oxford University Press 2024).

[10] On the ambiguity of human rights law, see I Kusche, "Possible Harms of Artificial Intelligence and the EU AI Act: Fundamental Rights and Risk" (2024) Journal of Risk Research 1. On the uncertainty inherent in the proportionality test that grounds human rights law, see L Enqvist and M Naarrijärvi, "Discretion, Automation, and Proportionality" in M Suski (ed), *The Rule of Law and Automated Decision-Making* (Springer 2023) 147.

[11] J Ziller, "The Council of Europe Framework Convention on Artificial Intelligence vs. the EU Regulation: Two Quite Different Legal Instruments" (CERIDAP 2/2024), 202, where it is explained that the CoE AI Convention will likely function as an interpretative document for the European Court of Human Rights.

[12] It is a standard interpretative practice of the European Court of Human Rights to use external sources, including other CoE treaties, when interpreting the European Convention on Human Rights. E Voeten, "Why Cite External Legal Sources? Theory and Evidence from the European Court of Human Right" in C Giorgetti and M Pollack (eds), *Beyond Fragmentation: Cross-Fertilization, Cooperation, and Competition among International Courts* (Cambridge University Press 2022) 162.

[13] See also Art. 27(2) of the CoE AI Convention that acknowledges the special relationship between the Convention and the EU AI Act. See also Council Decision (EU) 2024/2218 of 28 August 2024 on the signing, on behalf of the European Union, of the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law: "The Convention is to be implemented in the Union exclusively through Regulation (EU) 2024/1689 and other relevant Union acquis, where applicable."

[14] C McCrudden, "Pluralism of Human Rights Adjudication," in L Lazarus, C McCrudden, N Bowles (eds), *Reasoning Rights: Comparative Judicial Engagement* (Hart Publishing, 2014). See also M Delmas-Marty, *Towards a Truly Common Law: Europe as a Laboratory for Legal Pluralism* (translated by N Norberg) (Cambridge University Press).

[15] See also N Krisch, "The Open Architecture of European Human Rights Law" (2008) 71(2) The Modern Law Review 183.

[16] For clarifications of the role of Explanatory Reports for interpreting CoE treaties as related to the Vienna Convention on the Law of Treaties, (1969) Treaty Series, 1155, 331. see Section II.

might be interpreted.[17] The Feasibility Study prepared by the CoE Ad Hoc Committee on AI that proposed the adoption of a new CoE treaty as an option, will serve as a benchmark for identifying the intended objectives and evaluating the extent to which they have been achieved in light of the obligations adopted in the final text. The Feasibility Study can be considered to form part of the historical background of the Convention. It can be also considered as part of the circumstances of the conclusion of treaty, which in accordance with Article 32 of the Vienna Convention on the Law of Treaties, may be used as a supplementary means of treaty interpretation.[18] The drafts of the treaty that were published during the negotiations are also included in the analysis. These drafts form part of the preparatory work of the treaty that is also a supplementary means of treaty interpretation. These means that are based on key documents such as Feasibility Study, drafts and the Explanatory Report, are used for interpreting and better understanding the principles articulated in the CoE AI Convention, namely human dignity and individual autonomy, transparency and oversight, accountability, non-discrimination, data protection, reliability and risk management.

II. The context and the drafting process

Similarly to the AI Act, the CoE AI Convention has been inspired by, one the one hand, "the unprecedented opportunities" that AI might offer,[19] including by improving the efficiency of administrative procedures,[20] and by the risks that systems might pose to human rights, on the other.[21] In contrast to the AI Act, however, the CoE AI Framework Convention is not a product safety regime.[22] The CoE AI Convention has to be understood in light of the broader mandate of the organisation within which it has been adopted. In particular, the mandate of the CoE is focused on protection of human rights, democracy and rule of law.[23] This is also clearly reflected in Article 1(1) that stipulates that the treaty aims "to ensure that activities within the lifecycle of artificial intelligence systems are fully consistent with human rights, democracy and the rule of law."

The CoE's mandate shaped the origins of the treaty and allowed the involvement of various actors. In particular, the origins of the treaty can be traced back to May 2019 when the CoE Committee of Ministers recognised the need for examining "the feasibility and potential elements on the basis of multi-stakeholder consultations, of a legal framework for the development, design and application of artificial intelligence, based on the Council

---

[17] In the future, the work of the Conference of the Parties, composed of official representatives of the Parties to the Convention to determine the extent to which its provisions are being implemented, will also assist the interpretation. See Article 23 of the CoE AI Convention.

[18] Vienna Convention on the Law of Treaties of 23 May 1969, entered into force on 27 January 1980. United Nations, Treaty Series, vol. 1155, p. 331.

[19] See the Preamble of the treaty.

[20] CoE Ad hoc Committee on Artificial Intelligence (CAHAI) Feasibility Study, CAHAI(2020)23, 17 December 2020, para 20.

[21] As part of its efforts to reconcile innovation and regulation for the benefit of human rights, the CoE has also adopted Protocol to amend the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (CETS No. 223), 10 November 2018. See Section V.3.e below.

[22] The AI Act has been inspired by product safety regimes. See F Palmiotto, "The AI Act Roller Coaster: The Evolution of Fundamental Rights Protection in the Legislative Process and the Future of the Regulation" (2025) European Journal of Risk Regulation 1, 8; O Mir, "The AI Act from the Perspective of Administrative Law: Much Ado about Nothing?" (2024) European Journal of Risk Regulation 1, 3. Concerns have been expressed as to how AI-related risks, which concern ethical and fundamental rights issues, are addressed through harmonisation techniques which were developed to address health and safety concerns. See S de Vries et al, "Internal Market 3.0: The Old 'New Approach' for Harmonising AI Regulation" in G Robinson (et al, eds) "Future-proof Regulation and Enforcement for the Digitalised Age" (2023) 8 European Papers, Journal of Law and Integration 3.

[23] See first recital in the Preamble of the CoE AI Convention.

of Europe's standards on human rights, democracy and the rule of law.[24] In September 2019 the Committee of Ministers formally established the Ad Hoc Committee on Artificial Intelligence (CAHAI),[25] a body tasked to complete a feasibility study examining the possibility for the adoption of such a legal framework. According to the given mandate, this examination should be based on broad multi-stakeholder consultations, which could ensure the inclusion of diverse perspectives and concerns related to AI. In December 2020 CAHAI published its Feasibility Study,[26] where, as originally envisioned, multiple perspectives were presented and the foundations for a the new CoE treaty were formed. One option proposed in the Study was the adoption of a Framework Convention, as opposed to a Convention.[27] The latter imposes more concrete obligations and attributes rights to natural or legal persons.[28] A Framework Convention, on the other hand, is less prescriptive and detailed, it provides for "broad core principles and values" and leaves "a considerable margin of discretion for States as to how they implement the broader principles and values."[29] The advantages from the latter were three: first, detailed legal obligations for AI systems can be premature; second, States would be more willing to become bound; third, rigid regulation might obstruct innovation.[30] The implications of choosing a Framework Convention will become clearer in the forthcoming analysis.

Following the completion of the CAHAI's mandate, the Committee of Ministers established the Committee on Artificial Intelligence (CAI) with the task to "establish an international negotiation process and conduct work to finalise an appropriate legal framework on the development, design, use and decommissioning of artificial intelligence, based on the Council of Europe's standards on human rights, democracy and the rule of law."[31] It is important to understand the working process of CAI. This work was conducted in plenary sessions and drafting group sessions. As to the plenary sessions, there were ten such sessions.[32] The publicly available documents from these sessions offer no insights as to the positions of the different States involved.[33] The drafting group sessions were veiled with confidentiality.[34] Besides the four drafts that CAI decided to make public during the

---

[24] See 129th Session of the CoE Committee of Ministers (Helsinki, 16–17 May 2019), CM/Del/Dec(2019)1346/1.5, CM/Del/Dec(2019)1346/1.5

[25] See 1353rd meeting, 11 September 2019 of the CoE Committee of Ministers, CM/Del/Dec(2019)1353/1.5-app, CM/Del/Dec(2019)1353/1.5-app

[26] CAHAI(2020)23, Strasbourg 17 December 2020.

[27] A review of the complete list of the CoE treaties available here Full list – Treaty Office shows that the adoption of Framework Conventions is the exception.

[28] A comparison between the CoE AI Convention and Convention 108+ for the Protection of Individuals with Regard to the Processing of Personal Data, reveals the different levels of specificities. For example, the CoE AI Convention does not contain a provision similar to Art. 9 of Convention 108+. See Section V.3.e below.

[29] Feasibility Study, CAHAI(2020)23, para 134–9.

[30] Feasibility Study, CAHAI(2020)23, para 138.

[31] CM(2023)131-addfina, 1680ade00f

[32] Here references are given starting with the first one CAI(2022)06_rev, Strasbourg, 22 April 2022, 1680a6d912; CAI(2022)10, Strasbourg, 23 September 2022, 1680a83eaa; CAI(2023)03, Strasbourg, 13 January 2023, 1680a9cc4f; CAI(2023)05, Strasbourg, 3 February 2023, 1680aa182f; CAI(2023)11, Strasbourg, 21 April 2023, https://rm.coe.int/cai-2023-11-list-of-decisions/1680ab068f; CAI(2023)14, Strasbourg, 2 June 2023, 1680ab71a5; CAI(2023)23, Strasbourg, 26 October 2023, 1680ad13c6; CAI(2023)27, Strasbourg, 8 December 2023, 1680adc984; CAI(2024)04, Strasbourg, 26 January 2024, 1680ae51d8; CAI(2024)09, Strasbourg, 14 March 2024, 1680aef19f.

[33] StateWatch has published a document that contains some States' positions on the Zero Draft coe-artificial-intelligence-convention-compilation-of-comments-8-9-22.pdf

[34] These were informal meetings for resolving contentious issues. See CAI(2023)14, Strasbourg, 2 June 2023, 1680ab71a5: "Instruct the Secretariat to organise a series of informal meetings of the Drafting Group to explore possible solutions to some of the outstanding issues in the draft Framework Convention, during the period from June to December 2023."

process, "the preparatory works [of the CoE AI Convention], including the positions of different states, are not publicly accessible."[35]

Even in relationship to the drafts made public during the process, there were reservations added as to the value of these drafts.[36] There were also strong criticism by CAI of any breaches of the confidentiality of the process.[37]

As to the drafts made available during the drafting process, the Zero Draft was made public on 30 June 2022.[38] In February 2023 a Revised Zero Draft was published.[39] In July 2023 a Consolidated Working Draft of the Convention was published.[40] In December 2023 a draft before the final reading was published.[41] The final draft was adopted in March 2024 by the CAI[42] and later this was approved by the Committee of Ministers.[43]

The Zero Draft and the Revised Zero Draft (hereafter, where relevant, referred to collectively as the Zero Drafts) were largely similar in content. Substantive changes were introduced in the Consolidated Working Draft, released in July 2023. These changes highlight the key points of contention, which will be addressed in the analysis below in relation to specific provisions. It is worth noting that, although the drafting process – as envisioned under the CoE's mandate – involved a wide range of actors, including civil society stakeholders who provided input on the drafts,[44] the process was ultimately led and shaped by States. The EU was also involved.[45] Crucially, States that are not members of the CoE, such as USA, were also involved. Such a global reach was originally recommended by CAHAI.[46] Yet, it has been noted that the involvement of States, such as USA, diluted the safeguards and let to the exclusion of civil society.[47]

---

[35] The author of this article sent specific enquiries to CAI Secretariat. On 1 April 2025, the author received the following e mail "Please note that only the documents available on our website are open for public consultation. The preparatory works, including the positions of different states, are not publicly accessible." E-mail on file with the author.

[36] CAI(2023)05, Strasbourg, 3 February 2023, 1680aa182f: "Instruct the Secretariat to make the revised 'Zero Draft' public, with the information that this is a document prepared by the Chair and the Secretariat and does not reflect the final outcome of negotiations in the Committee."

[37] CAI(2023)05, Strasbourg, 3 February 2023, 1680aa182f: "Take note of the opening remarks of Mr Thomas Schneider, Chair of the CAI. In his remarks, the Chair referred to the recent breaches of confidentiality of meetings and documents of the CAI, including references having been made to specific positions of Delegations. *He emphasized that such actions were unacceptable.* He reiterated the requirements under Resolution CM/Res (2021)3 'on intergovernmental committees and subordinate bodies, their terms of references and working methods,' adopted by the Committee of Ministers on 12 May 2021. The Chair further explained that the need for confidentiality of meetings and documents of the CAI was not to curtail transparency, but because international negotiations require the ability for Delegates to make statements or take positions without being quoted in public. Transparency and inclusivity are important aspects of modern policy making and in the case of the CAI, they can and will be ensured in appropriate ways." (emphasis added).

[38] The Zero Draft can be accessed here coe-artificial-intelligence-convention-zero-draft-30-6-22.pdf.

[39] The Revised Zero Draft can be accessed here 1680aa193f.

[40] 1680abde66.

[41] Convention on AI and human rights (draft December 2023) | Digital Watch Observatory.

[42] CM(2024)52-prov1 available 1680aee411.

[43] CM/Del/Dec(2024)1497/10.1, 1497th meeting, 30 April 2024, CM/Del/Dec(2024)1497/10.1.

[44] See, for example, the involvement of the European Network of National Human Rights Institutions here Draft Convention on AI, Human Rights, Democracy and Rule of Law finalised: ENNHRI raises concerns – ENNHRI and the following civil society statement CSO-COE-Statement_07042023_Website.pdf.

[45] See Council Decision (EU) 2022/2349 of 21 November 2022 authorising the opening of negotiations on behalf of the EU for a CoE convention on artificial intelligence, human rights, democracy and the rule of law.

[46] 'The CAHAI therefore recommended that the instrument, "though obviously based on Council of Europe standards, be drafted in such a way that it facilitates accession by States outside of the region that share the aforementioned standards."

[47] See the Civil Society Statement CSO-COE-Statement_07042023_Website.pdf See also here US obtains exclusion of NGOs from drafting AI treaty – Euractiv.

The precise influence of individual States – including non-Member States of the CoE – on the drafting process is difficult to document. State positions regarding the meaning of specific treaty provisions have not been made public, which complicates interpretative efforts. It is important to emphasise, however, that these features of the drafting process – including its confidential nature – are not unique to the CoE AI Convention. Rather, they reflect broader practices within the CoE's treaty-making framework. As such, the AI Convention should be understood within this institutional context.

For this understanding, it is relevant to observe that as early as the 1960s, a proposal to make the *travaux préparatoires* of CoE conventions – and, by extension, the negotiating positions of States – public was rejected.[48] Yet, at the same time, it was also agreed that "it would be appropriate to publish a report which, without revealing the attitudes of the various experts of governmental delegations during the proceedings, would be of a nature to facilitate the application of their provisions."[49] This is how it was decided that the committee of experts (in relationship to the CoE AI Convention, this was CAI) that prepared the draft convention, would also issue an explanatory report.[50]

At this point, an issue that arises concerns the role of the Explanatory Report in the interpretation of the treaty. The report does not constitute an authoritative interpretation of the text.[51] This is reflected in the second sentence of the CoE AI Convention Explanatory Report: "[t]he text of the explanatory report submitted to the Committee of Ministers of the Council of Europe" does not constitute an instrument providing an authoritative interpretation of the text of the Framework Convention although it may facilitate the understanding of it's provisions." Since the text of the treaty and the report are negotiated and adopted simultaneously, the report can be viewed as a supplementary means of treaty interpretation in the sense of Article 32 of the Vienna Convention on the Law of Treaties.[52] The Report can be also considered as part of the context in which the meaning of certain provisions is to be ascertained in the sense of Article 31(1) of the Vienna Convention. In the analysis that follows, the Explanatory Report, along with the available Drafts and the Feasibility Study – which together shed light on "the preparatory work of the treaty and the circumstances of its conclusion",[53] will serve as important sources for interpreting the meaning treaty's provisions. A crucial first step in this process is to clarify the treaty's personal scope.

III. Scope: public authorities and private actors

1. Negotiating the scope

As the publicly available Drafts show, the personal scope of the Convention was an object of major changes. In particular, the Zero Drafts contained the concepts of "AI provider," "AI user" and "AI subject." These were removed with the Consolidated Working Draft. The references to "AI provider" and "AI user" allowed the Zero Drafts to directly regulate private actors, which was also reflected in the proposed personal scope of the Convention.

[48] See CoE Parliamentary Assembly Recommendation 417 (1965) of 25 January 1965 for publication of the preparatory works. However, "[i]n view of the confidentiality of the experts' working documents, the Committee of Ministers could not accept this proposal." Jörg Polakiewicz, *Treaty-making in the Council of Europe* (Council of Europe Publishing, 1999), 26.

[49] Reply by the Committee of Ministers to Recommendation 417 (1965), adopted during the 145th meeting of the Ministers' Deputies in October 1965, cited in Jörg Polakiewicz, *Treaty-making in the Council of Europe*, (Council of Europe Publishing, 1999), 26.

[50] J Polakiewicz, *Treaty-making in the Council of Europe*, (Council of Europe Publishing, 1999), 26. The CoE Committee of Ministers authorises the publication of the Explanatory Report.

[51] J Polakiewicz, note 50, 26.

[52] Ibid, 27.

[53] See Article 32, Vienna Convention on the Law of Treaties.

In particular, the proposed scope contained the clarification that the Convention applies "regardless of whether these activities [design, development and application of AI systems] are undertaken by public or private actors."[54] This direct regulation of private actors was removed with the Consolidated Working Draft. This removal is retained in the final version of the text.

The concept of "AI subject" allowed the Zero Drafts to contain multiple references to concrete rights of the "AI subjects."[55] Relevant examples include: the right of access to the relevant records;[56] the right to human review;[57] and the right to know that one is interacting with an AI system.[58] Such concrete formulations of rights were removed with the Consolidated Working Draft.[59] This removal is retained in the final version of the text.

2. Understanding the scope

To better understand the scope of the treaty, the sections that follow clarity how the usage of AI systems by public authorities (Section III.2.a), by private actors whose conduct is attributable to the State (Section III.2.b.) and by private actors whose conduct is *not* attributable to the State (Section III.2.c.), is regulated.

a. Public authorities

CoE AI Convention regulates the conduct of States, not directly of private actors, which distinguishes it from the EU AI Act that regulates directly *both* private and public providers and deployers of AI systems.[60] As other CoE conventions that are international treaties by which States agree to adopt obligations under international law, the treaty applies to its State Parties. This implies that it applies to public authorities. Article 3(1)(a) of the treaty stipulates that "Each Party shall apply this Convention to the activities within the lifecycle of artificial intelligence systems undertaken by public authorities, [. . .]". This follows the rules from general international law. In particular, it corresponds to Article 4 of the International Law Commission (ILC) Articles on Responsibility of States for Internationally Wrongful Acts, which provides that "[t]he conduct of any State organ shall be considered an act of that State under international law, whether the organ exercises legislation, executive, judicial or any other functions, whatever position it holds in the organisation of

---

[54] Art. 4(1) Zero Draft CAI(2022)07 Restricted, 30 June 2022; Art. 4(2), Revised Zero Draft, CAI(2023)01, 6 January 2023.

[55] Here I make a distinction between, on the one hand, general references to human rights, human rights obligations or impact on human rights (that are multiple in the Consolidated Working Draft and in the final version of the CoE AI Convention), and, on the other, conferral of concrete rights specifically relevant to the application of AI systems. Such concrete rights include, for example, the right to know that one interacts with an AI system or the right to choose not to interact.

[56] Art. 7(1) Zero Draft.

[57] Art. 7(2) Zero Draft; Article 20(1) Revised Zero Draft.

[58] Art. 7(4) Zero Draft. Article 20(2) Revised Zero Draft.

[59] Yet, see Art. 14(2) of the Consolidated Working Draft that stipulated that "*any person has the right to know* that one is interacting with an AI system rather than with a human unless obvious from the circumstances and context of use and, where appropriate, shall provide for the option of interacting with a human in addition to, or instead of, such system." (emphasis added) Compare this proposal in the Consolidated Working Draft with the final version as reflected in Article 15(2) of the AI Convention that contains no reference to a *right* and to an option to choose not to interact with AI.

[60] Similarly to private entities and individuals, public authorities can be providers when they *develop* an AI system and put it into service (Art. 2(3), AI Act). Similarly to private entities and individuals, public authorities can be also deployers of AI systems when they *use* the systems (Art. 2(4), AI Act.). See O Mir, "The AI Act from the Perspective of Administrative Law: Much Ado about Nothing?" (2024) European Journal of Risk Regulation 1, 9.

the State, and whatever its character as an organ of the central Government or of a territorial unit of the State."[61] The clarifications included in the Explanatory Report to the CoE AI Convention follow the same principle: "the Drafters' shared understanding is that the term 'public authority' means any entity of public law of any kind or any level (including supranational, State, regional, provincial, municipal, and independent public entity) [. . .]."[62]

b. *Private actors whose conduct is attributable to the State*

Article 3(1)(a) of the treaty, however, adds that the Parties shall apply also the Convention to the activities undertaken by "private actors acting on their [public authorities'] behalf." The Explanatory Report clarifies that this means that the Convention imposes obligations "in regard to activities for which public authorities *delegate* their responsibilities to private actors *or direct them to act,* such as activities by private actors operating pursuant to a contract with a public authority or other private provision of public service, as well as public procurement and contracting."[63] This also follows the rules from general international law. In particular, it corresponds to Articles 5 and 8 of the ILC Articles on Responsibility of States for Internationally Wrongful Acts.

Articles 5 of the ILC Articles on Responsibility of States for Internationally Wrongful Acts provides that the activities of entities exercising elements of governmental authority *are considered activities of the State*, provided that these entities are empowered by the law of that State to exercise elements of the governmental authority and that they are acting "in that capacity in the particular instance." While the ILC Articles, Article 3(1)(a) of the CoE AI Convention, and the Explanatory Report employ different terminology and concepts – creating some ambiguity – the key point is clear: when public authorities delegate tasks to private actors who then deploy AI systems, the Convention remains applicable. In such cases, it is the public authorities that bear ultimate responsibility, as the use of AI systems is attributed to the State.

As to Article 8 of the ILC Articles on Responsibility of States for Internationally Wrongful Acts, it provides that activities of private entities are considered activities of the State if these private entities act on the instructions of, or under the direction or control of the State in carrying out these activities. This aligns with the language in Article 3(1)(a) of the CoE AI Convention that refers to private entities acting on behalf of the State. It also aligns with the framing in the Explanatory Report, which refers to situations where public authorities direct private actors to undertake activities. If these activities include the usage of AI systems, the Convention applies.

In simple terms, all of the above means that there is a group of private entities whose conduct is attributable to the State because of delegation or control that implies the exercise of public powers. Any activities within the lifecycle of AI systems that these entities perform are therefore activities by the State Parties and all the obligations formulated in the CoE AI Convention apply. In this sense, the activities of these private entities are considered activities of the State. A possible contentious question here is the interpretation of the meaning of "acting on behalf" of the State authorities,[64] "prerogatives of official authority,"[65] and "delegation of their [public authorities']

---

[61] Draft Articles on Responsibility of States for International Wrongful Acts with Commentaries, *Yearbook of International Law Commission*, 2001, Vol.II, Part Two.

[62] Explanatory Report to the CoE AI Convention, para 27.

[63] Ibid, para 28 (emphasis added).

[64] The formulation used in Art. 3(1)(a), CoE AI Convention. Explanatory Report to the CoE AI Convention, para 28.

[65] Explanatory Report to the CoE AI Convention, para 27.

responsibilities to private actors."[66] The contention relates more generally to the meaning of public authority and of public powers.[67]

c. Private actors whose conduct is not attributable to the State

Article 3(1)(b) of the CoE AI Convention addresses a *second group of private entities*. These are entities *not* covered by Article 3(1)(a) of the treaty and, in this sense, these are entities whose conduct is not attributable to the State since they do not exercise public powers. The conduct of these entities is not directly an object of regulation by the treaty. It is rather the State Parties that "shall address risks and impacts arising from" these private entities' activities within the lifecycle of AI systems. It is the States that adopt the obligation under the treaty to "address the risks and impacts."

States have two choices for how to fulfil this obligation. First, States can choose to apply the principles and obligations of the CoE AI Convention to private entities' activities. This implies that States can choose to regulate at national level private actors by imposing requirements upon the latter similar to how the obligations in the treaty are framed. This could imply adopting legislation of, for example, private law nature that refers to concepts such as human dignity, individual autonomy, reliability, human rights etc., as used in the text of the treaty, to regulate private actors.

Second, States can choose to "take other appropriate measures to fulfil the obligation" to address risks and impacts.[68] This second choice seems to imply that States enjoy wider discretion in regulating risks and impacts at the national level.

Ultimately, though, the difference between the first and the second option is hard to clearly distinguish. The Explanatory Report says generally in respect to Article 3(1)(b) of the CoE AI Convention that it requires "the adoption or maintaining of appropriate legislative, administrative or other measures to give effect to this provision." The Explanatory Report also adds that

> [. . .], the obligation does not necessarily require additional legislation and Parties may make use of other appropriate measures, including administrative and voluntary measures. So while the obligation is binding and all Parties should comply with it, *the nature of the measures taken by the Parties could vary.*[69]

Overall, then, private actors (that fall within the second group) are regulated indirectly "by virtue of the rights granted to, and obligations assumed by states" under the CoE AI Convention.[70] The treaty therefore can indirectly address private actors, which is very similar to how the European Convention on Human Rights (hereinafter the ECHR) can

---

[66] Explanatory Report to the CoE AI Convention, para 28.

[67] It has been well established in the ECtHR's case law that States are not absolved of their human rights law obligations by delegating certain services to private bodies (see, e.g., *Costello-Roberts v United Kingdom* App no 13134/ 87 (ECHR, 25 March 1993); *Dodov v Bulgaria* App no 59548/00 (ECHR, 17 January 2008) para 80; *Storck v Germany* no 61603/00 (ECHR, 16 June 2005) para 103; *O'Keeffe v Ireland* [GC] App no 35810/09 (ECHR, 28 January 2014) para150. This is especially so when these services are considered to be public services. Yet, the meaning of public services and relatedly, the meaning of public powers, is not unambiguous. One way will be to say that public functions are the ones that the State has historically performed. For a more in-depth discussion, see J Thomas, *Public Rights, Private Relations* (Oxford University Press 2015) 41.

[68] States are obliged to make a declaration which of these two options they have chosen. Norway has made a declaration that it has made the first choice.

[69] Explanatory Report to the CoE AI Convention, para 29 (emphasis added).

[70] CoE Ad hoc Committee on Artificial Intelligence (CAHAI) Feasibility Study, CAHAI(2020)23, 17 December 2020, para 91.

indirectly regulate private actors.[71] It can therefore be argued that the CoE AI Convention imposes positive obligations on States to regulate private actors.[72] Such positive obligations include preventive measures meant to *ex ante* protect individuals from harm or risk of harm.[73]

IV. The risk-based approach

The obligations imposed by the treaty are shaped by a risk-based approach. According to Article 1(2) of the CoE AI Convention, the measures meant to regulate AI systems "shall be graduated and differentiated as may be necessary in view of the severity and probability of the occurrence of adverse impacts on human rights, democracy and the rule of law throughout the lifecycle of artificial intelligence systems."[74] The phrase "adverse impacts on human rights" is noteworthy and raises the question about the difference between risks and "adverse impacts."[75] To make things even more confusing, the Convention also refers to "potential to interfere with human rights."[76] Potentiality seems to more closely align with the notion of risk. Ultimately, analytical confusion permeates the text of the Convention as to the role of risk in the conceptualisation of harm and the measures meant to be prevent and address harm (or risk of harm). As Section V.2. will show, this confusion existed from the very beginning of the drafting process and undermines the clarity of the obligations undertaken by States.

What is relevant at this juncture is that generally the regulatory model can be characterised as one based on risk. Therefore, similarly to the EU AI Act,[77] the CoE AI Convention has a risk-based approach: the measures "need to be tailored to the level of risk posed by an artificial intelligence system within the specific spheres, activities and contexts."[78] The State Parties have discretion how to balance any competing interests in each sphere, activity or context.[79] The Explanatory Report notes, though, that the public sector sphere, might have some specificities that need to be taken into account.[80] Examples of these specificities are not provided; rather, the Explanatory Report gives examples of

---

[71] Vladislava Stoyanova, *Positive Obligations under the European Convention on Human Rights. Within and Beyond Boundaries* (Oxford University Press 2023); Claire Loven, *Fundamental Rights Violations by Private Actors and the Procedure before the European Court of Human Rights. A Study of Verticalised Cases* (Intersentia 2022).

[72] These positive obligations are without prejudice to other obligations (positive or negative) under human rights treaties. See third paragraph in Article 3(1)(b) of the CoE AI Convention.

[73] See, generally, V Stoyanova, *Positive Obligations under the European Convention on Human Rights. Within and Beyond Boundaries* (Oxford University Press 2023). Many scholars have argued that it is precisely positive obligations in human rights law that demand from States to better regulated AI systems so that harms can be prevented. See L McGregor, D Murray and V Ng, "International Human Rights Law as a Framework for Algorithmic Accountability" (2019) 68 International & Comparative Law Quarterly 309; L Lane, "Clarifying Human Right Standards Through Artificial Intelligence Initiatives" (2022) 71(4) International and Comparative Law Quarterly 915.

[74] The Explanatory Report to the treaty is useful in better understanding what is meant by the "lifecycle" of the system. In particular, this lifecycle includes (1) planning and design, (2) data collection and processing, (3) development of artificial intelligence systems, including model building and/or fine-tuning existing models for specific tasks, (4) testing, verification and validation, (5) supply/making the systems available for use, (6) deployment, (7) operation and monitoring, and (8) retirement. See Explanatory Report, para 15.

[75] See also Arts. 9 and13 of the CoE AI Convention that refer to "adverse impacts on human rights."

[76] See Arts. 3(1) and14(2)(a) CoE AI Convention.

[77] T Mahler, 'Between Risk Management and Proportionality: The Risk-Based Approach in the EU Artificial Intelligence Act Proposal' in L Colonna and S Greenstein (ed), *Law in the Era of Artificial Intelligence.* Nordic Yearbook of Law and Informatics 247.

[78] Explanatory Report, para 17.

[79] Ibid.

[80] Ibid.

public sector areas (i.e., law enforcement, migration, border control, asylum and the judiciary), where the existence of specificities should be assumed. It also highlights that in certain spheres, there are "power asymmetries," where certain groups within the society are in a weaker and vulnerable position, which needs to be taken into consideration.[81]

Contrary to the EU AI Act, the CoE AI Convention does not create different levels of risks (e.g., unacceptable risks, high risks, minimum risks).[82] It does not therefore differentiate AI systems based on the risk that they might pose. The Zero Drafts did envision different levels of risk and measures tailored to the levels. Three levels were regulated in the Zero Draft: first, "significant levels of risk";[83] second, "unacceptable levels of risk" that were intended to lead to full or partial ban only after the unacceptable risks have been identified and there were "no other measures available for mitigating" them;[84] third, unacceptable levels of risk that were intended to lead to direct bans of such systems.[85] These levels were removed very early in the drafting process.[86] Due to the confidentiality of the negotiating process, the precise reasons cannot be documented.

Since the CoE AI Convention does not differentiate levels of risk, the obligations imposed by the treaty are not tailored with reference to the level of risk a system might be accepted to pose. Neither is the regulation by the treaty tailored to specific sectors (such as asylum, immigration and border control).[87] Rather Chapter II of the Convention imposes general obligations that apply to all AI systems that fit within the Article 2 definition.[88] As already clarified, the addressees of these obligations are States.

The absence of defined risk levels in the CoE AI Convention is a key point of contrast with the EU AI Act and warrants a more elaborate consideration here. While a detailed examination of the AI Act's differentiated regulation of risk levels lies beyond the scope of this analysis, it is important to acknowledge the significance of these levels – particularly given their inherent ambiguities and the potential they create for regulatory exclusions. This observation, in turn, provides a useful point of comparison with the CoE AI Convention, which also adopts a risk-based regulatory approach but does so without explicitly categorising risk levels and tailoring obligations accordingly.[89] One might initially assume that the Convention provides a lower level of protection, as it omits a high-risk category present in the EU AI Act. Nevertheless, as the subsequent analysis will illustrate, this assumption is overly simplistic given the avenues available for avoiding the high-risk designation under the EU AI Act.

---

[81] According to the Explanatory Report, para 18, these spheres included "distribution of social welfare benefits, decisions on the creditworthiness of potential clients, staff recruitment and retention processes, criminal justice procedures, immigration, asylum procedures and border control, policing, and targeted advertising and algorithmic content selection."

[82] See E Leinarte, "The Classification of High-Risk AI Systems Under the EU Artificial Intelligence Act" (2024) 3 Journal of AI Law and Regulation 262; M Almada and N Petit, "The EU AI Act: Between the rock of product safety and the hard place of fundamental rights" (2025) 62(1) Common Market Law Review 85.

[83] Art. 12, Zero Draft CAI(2022)07, 30 June 2022.

[84] Art. 13, Zero Draft CAI(2022)07, 30 June 2022.

[85] Art. 14, Zero Draft CAI(2022)07, 30 June 2022.

[86] See the Revised Zero Draft CAI(2023)01, 6 January 2023.

[87] See Annex III of the EU AI Act.

[88] See the exclusion of AI systems related to the protection of States' national security interests and national defence, as stipulated in Art. 3(2) and (4) CoE AI Convention. "National security" is distinguished from "public security." See Explanatory report, para 32.

[89] The comparison presented here has certain limitations that should be openly acknowledged. The analysis in this section focuses specifically on Art. 6(3) of the EU AI Act, which sets out the classification rules for high-risk AI systems. It does not attempt to provide a comprehensive examination of the Act's overall risk-based approach. For example, practices prohibited under Art. 5 of the AI Act – a provision that does not allow for general derogations – are not addressed.

Within the logic of the AI Act, risk levels play a central role. The Act has been described as aiming primarily to regulate those AI systems classified as posing high or unacceptable risk. To expose the ambiguities surrounding the risk classifications and the resulting regulatory uncertainties, particular attention can be given to the designation of "high-risk" AI systems. The AI Act does not define "high-risk" in general. Article 6 with its reference to Annex III of the Act indicates areas where the usage of AI systems can be assumed to be high-risk as a starting point.[90] Yet, a system listed in Annex III is *not* automatically assumed to be high-risk. More specifically, a system used, for example, in the examination of asylum claims is *not assumed to be a high-risk* by the mere fact that the system is used in a sensitive area, such as asylum, by public authorities that eventually take crucial decisions affecting human beings. Pursuant to Article 6(3) of the AI Act an additional threshold needs to be passed for such a system to be classified as a "high-risk."[91]

This provision stipulates that

> an AI system referred to in Annex III shall *not* be considered to be a high-risk where it does not pose a significant risk of harm to the health, safety or fundamental rights of natural persons, including by not materially influencing the outcome of decision making (emphasis added).[92]

A contrario, an AI system assisting competent authorities in the examination of, for example, asylum claims, is classified as a high-risk system *only* if it poses "a significant risk of harm to the health, safety or fundamental rights of natural persons." Here it is pertinent to clarify that it is the system *itself* that has to pose "a significant risk," *not* the final decision taken by the public authority. The final decision (e.g., rejection of an asylum claim) might pose a risk of harm to the health, safety or fundamental rights. Yet, this is not what matters. What matters for the classification of the AI system as a "high-risk" is that the system *itself* when used to assist a decision poses this type of risk.

This raises difficult causation questions.[93] More specifically, it needs to be possible to demonstrate the *causal link* between the AI system and the risk of harm to health, safety and fundamental rights of persons. If the causal link between the AI system and the risk is shown, the system can be classified as a "high-risk" system. The question as to the relevant standard of causation is not at all addressed in the AI Act. The question as to how to

---

[90] Notably, the AI systems included in Annex III may increase or decrease, since new systems can be added or systems can be removed from the list. The addition of new examples of high-risk systems is meant to ensure flexibility as to how the technology is regulated given that it might develop very rapidly. Art. 97 AI Act allows the Commission to adopt delegated acts under Art. 290 TFEU to amend some of its provisions (the scrutiny by the Council and the European Parliament is still required). Annex III can be also an object of such amendments whose areas the Commission can extend (only in the framework of the eight existing areas) in accordance with the substantive criteria set out in Art. 7 AI Act. The Commission can also change the exception of Art. 6(3) AI Act.

[91] This threshold does not apply to AI systems that profile natural persons. Art. 6(3) third sentence stipulates that "an AI system referred to in Annex III shall always be considered to be high-risk where the AI system performs profiling of natural persons."

[92] The objective of this provision is to reduce the regulatory burden of the AI Act: "the significance of the output of the AI system in relation to the decision or action taken by a human, as well as the immediacy of the effect, should also be taken into account when classifying AI systems as high risk." See Second Presidency Compromise text (11124/22, 15.7.2022) at page 4, AIA-CZ-1st-Proposal-15-July.pdf For a more detailed analysis of Art. 6 of the AI Act, see also Emilija Leinarte, 'The Classification of High-Risk AI Systems Under the EU Artificial Intelligence Act' (2024) 3 Journal of AI Law and Regulation 262, where the legislative process of Art. 6(3) of the AI Act is described and it is explained how the co-legislators "differed on both the threshold for risk and methodology for assessing it."

[93] See, generally, S Steel, "Legal Causation and AI" in E Lim and P Morgan (eds), *The Cambridge Handbook of Private Law and Artificial Intelligence* (Cambridge University Press, 2024) 189.

measure the level of significance of the risk is also left open. Notably, Article 6(3) of the AI Act does not simply refer to risk, but only to "*significant* risk of harm."

The usage of the concept of risk here is pertinent to highlight. It is not a causal link between the usage of the AI system and harm that needs to be shown, rather a causal link between the usage of the AI system and the *risk* of harm. This makes the causal inquiry easier since risk of harm is only a hypothetical. In then follows that since a causal link to risk is easier to show, the classification of a system as a "high-risk" system might be easier. Yet, as mentioned above, the risk needs to be "significant." The severity threshold that is implied in the concept of "significant risk" might offset the above-mentioned relative easiness of demonstrating risk instead of demonstrating harm.

Even if only risk of harm (as opposed to harm) needs to be demonstrated, we need to have some understanding of the harm. The question then that arises is the following: Harm to what? Article 6(3) of the AI Act specifies "harm to the health, safety or fundamental rights of natural persons." It might be difficult to imagine circumstances where health and safety are harmed, while fundamental rights are not harmed simultaneously. This makes the concurrent enumeration of these three concepts (i.e., health, safety or fundamental rights) difficult to comprehend.

It can nevertheless become comprehensible with the clarification that harm to safety and health, even if it infringes upon the important interests protected by human rights law (e.g., life, private and family life), does not necessary amount to violation of fundamental rights.[94] The reason is that even if the right to, for example, private life is infringed, this infringement can be proportionate and therefore not in violation of human rights law. This implies an inclusion of a proportionality analysis in the assessment whether the AI system poses a risk of harm. Such proportionality analysis implies highly complex and context-dependent reasoning.[95] At the end of the day therefore, the inclusion of fundamental rights *in the threshold* for classifying systems as high-risk, is *not* very helpful.

The inclusion of significant risk to fundamental rights in addition to significant risk to health and safety in Article 6(3) of the AI Act, can be interpreted to the effect that since fundamental rights do not cover only health and safety, the types of possible harms are wider. In other words, possible harms can include harms to privacy, family life, religious believes, procedural rights, etc. This expansion can be viewed in a positive light. Yet, as noted above, given that harms to interests such as privacy, family life etc. can be proportionate and thus *not* in violation of human rights, the expansion might not be very meaningful.

Article 6(3) of the AI Act stipulates that an AI system "does not pose a significant risk of harm to the health, safety or fundamental rights of natural persons, *including by not materially influencing the outcome of decision making.*" The role of the addition, reproduced here in italics, requires closer examination. It could be interpreted as introducing the assumption that if an AI system does not materially influence the outcome – such as a decision – it should be presumed not to pose a significant risk, and therefore would not fall under the classification of a "high-risk" system.

When does an AI system "materially" influence a decision (i.e., the outcome of the decision-making)? Article 6(3) of the AI Act introduces *assumptions* when this is *not* the case. Under the following four conditions, an AI system is assumed *not* to materially influence the outcome and therefore *not* to be a high-risk system. These are when the system is "intended to perform a narrow procedural task," "to improve the result of a previously completed human activity," "to detect decision-making patterns or deviations from prior decision-making patterns and is not meant to replace or influence the

---

[94] I use the terms "fundamental rights" and "human rights" interchangeably.

[95] L Enqvist and M Naarrijärvi, "Discretion, Automation, and Proportionality" in M Suski (eds) *The Rule of Law and Automated Decision-Making* (Springer 2023) 147.

previously completed human assessment, without proper human review," and "to perform a preparatory task to an assessment relevant for the purposes of the use cases listed in Annex III."[96] These four conditions are given as alternatives.

The framing of the four conditions, is open to interpretation. Whether a system performs a *narrow* procedural task can be a contentious question. The meaning of "narrow procedural task" is ambiguous. The same applies to the actual meaning of improving the result, not influencing and having only a preparatory task. This interpretational ambiguity can be addressed in the future via guidelines by the Commission.[97] The point here is that the conclusion that a system, for example, performs narrow procedural task, might not be that straightforward and might depend on the context.[98]

To conclude, the AI Act introduces ambiguous exceptions to the classification of systems as high-risk, which ultimately appears to undermine its regulatory approach with categorisations of risk levels and the tailoring of the obligations to the levels. The drafters of the CoE AI Convention have avoided this, thereby sidestepping the above-mentioned difficulties (causation, proportionality review, contextualisation etc.) in the initial operation of classifying systems into risk levels. This can be also related to its nature as an international treaty whose provisions are more abstract and general.

Although not related to the classification of risk levels, the previously noted difficulties – regarding the role of human rights, the causal links between harm (or the risk of harm) to important interests and the use of AI systems, proportionality and contextualisation – also arise for interpreting the obligations imposed by the CoE AI Convention, as the next section will explore.

V. The obligations imposed upon states

This section explains the relevant obligations by proceeding in three steps. First, it addresses the treaty's starting point: the express intention not to create *new* human rights obligations. Second, while the protection of human rights was initially identified as a gap that the CoE AI Convention was meant to fill, the lack of specification of the *arguably existing* human rights obligations raises doubts as to whether this gap has in fact been addressed. Finally, although the treaty does not concretise such obligations, it enshrines a number of principles specifically relevant to the operation of AI systems. These principles – human dignity and individual autonomy, transparency and oversight, accountability, non-discrimination, data protection, reliability, and risk management – are deliberately framed at a high level of generality but can nonetheless provide interpretative guidance for the future development of more concrete human rights standards and obligations.

---

[96] See also Art. 6(6) AI Act that empowers the Commission to add new or to modify these four conditions upon evidence that AI systems listed in Annex III do not pose "a significant risk of harm to health, safety or fundamental rights of natural persons." This can be interpreted as a possibility of deregulation. See also Art. 6(7) of the Act that can be viewed as a possibility for further regulation. It obliges the Commission to amend these four conditions by deleting any of them "where there is concrete and reliable evidence that this is necessary to maintain the level of protection of health, safety and fundamental rights."

[97] See Art. 6(5) AI Act that stipulates that the Commission no later than 2 February 2026 has to issue guidance including practical examples of AI systems that are high-risk and not high-risk.

[98] See also F Palmiotto, "When Is a Decision Automated? A Taxonomy for a Fundamental Rights Analysis" (2024) German Law Journal 1, 11, where it is noted "Worryingly, the new version of the AI Act seems to suggest that fundamental rights are unaffected when AI systems do not have a prevalent role in decision-making."

1. Starting point and gaps meant to be addressed

To better understand the obligations imposed upon States by the CoE AI Convention, it is pertinent to underscore its starting point. The treaty does not aim to establish new human rights law obligations. This is made very clear in its Explanatory Report:

> no provision of this Framework Convention is intended to *create new human rights or human rights obligations* or undermine the scope and content of the existing applicable protections, but rather, by setting out various legally binding obligations contained in its Chapters II to VI, *to facilitate* the effective implementation of the applicable human rights obligations of each Party in the context of the new challenges raised by artificial intelligence.[99]

The text of Article 4 of the CoE AI Convention reflects this starting point. This provision imposes an obligation upon each State Party to "adopt and maintain measures to ensure that the activities within the lifecycle of AI systems are consistent with obligations to protect human rights, as enshrined in application international law and its domestic law." This text does not add much specificity. It does not add much to the uncertainty regarding the answer to the question how human rights law limits and regulates AI systems.[100]

To better understand the role of the CoE AI Convention and its possible added value, it is useful to go back to the Feasibility Study prepared prior to the adoption of the Convention. The Study highlights four "legal gaps": (1) specification/concretisation of obligations; (2) identification of concrete principles specifically relevant to AI systems; (3) responding to systemic harms; and (4) responding to cross-border/transboundary nature of the harm. The following two subsections will concentrate on the first and second gaps,[101] given that most of the provisions adopted in the treaty pertain to them.[102]

---

[99] Explanatory report, para 13 (emphasis added).

[100] There are generally two lines of contributions: one argues that human rights law is not helpful, and another one argues that human rights law is well-equipped to face harms that might be caused by AI systems. For examples of the first one see S A Teo, "How Artificial Intelligence Systems Challenge the Conceptual Foundations of the Human Rights Legal Framework" (2022) 40(1) Nordic Journal of Human Rights 216; in-Y Liu, "AI Challenges and the Inadequacy of Human Rights Protections" (2021) 40 Criminal Justice Ethics 2; H-Y Liu, "The Digital Disruption of Human Rights Foundations" in M Susi (ed), *Human Rights, Digital Society and the Law: A Research Companion* (Routledge 2019). For scholarship that fits within the second line (i.e., human rights law is useful), see L McGregor, D Murray and V Ng, "International Human Rights Law as a Framework for Algorithmic Accountability" (2019) 68 International and Comparative Law Quarterly 309; E Donahoe and M MacDuffee Metzger, "Artificial Intelligence and Human Rights" (2019) 30 Journal of Democracy 115.

F Palmiotto, "When Is a Decision Automated? A Taxonomy for a Fundamental Rights Analysis" (2024) German Law Journal 1; S Demkova, "The EU's Artificial Intelligence Laboratory and Fundamental Rights" in M Fink (ed), *Redressing Fundamental Rights Violations by the EU: The Promise of the 'Complete System of Remedies* (Cambridge University Press, 2025)* 391.

[101] See Art. 5 of the CoE AI Convention for the third gap. As to the fourth gap, see Arts. 23 and 25.

[102] The third legal gap identified in the Feasibility Study and meant to be addressed by the CoE AI Convention, concerns the wider social impact from the deployment of AI systems (see CoE Ad hoc Committee on Artificial Intelligence (CAHAI) Feasibility Study, CAHAI(2020)23, 17 December 2020, para 86). This impact transcends human rights law that is focused on the individual and on more specific harms that affect specific identifiable individuals. This impact relates to broader questions about the operation of liberal democracies, including, for example, how AI systems might interfere with electoral processes and democratic institutions. The fourth legal gap identified in the Feasibility Study concerns the "lack of common norms at international level," which is an obstacle for the trade of AI systems and for mutual trust. CoE Ad hoc Committee on Artificial Intelligence (CAHAI) Feasibility Study, CAHAI(2020)23, 17 December 2020, para 88.

2. Protection of human rights

As to the first legal gap, the Feasibility Study notes that "the rights and obligations formulated in existing legal instruments tend to be articulated broadly or generally, which is not problematic as such, yet can in some instances raise interpretation difficulties in the context of AI."[103] Concretisation of the obligations corresponding to the rights is therefore proposed. The Feasibility Study suggests that this can be attained by "specifying more concretely what falls under a broader human right and how it could be invoked by those subjected to AI systems."[104]

Here it is argued that CoE AI Convention does not really achieve such a concretisation. Article 4 of the treaty simply refers to the obligation upon States to protect human rights. The Explanatory Report to the Convention does little to clarify – and in some respects further obscures – the nature of Article 4. Namely, paragraph 38 from the Explanatory Report states that

> [], parties are free to choose the ways and means of implementing their international legal obligations, provided that the result is in conformity with those obligations. *This is an obligation of result and not an obligation of means.* In this respect, the principle of subsidiarity is essential, putting upon the Parties the primary responsibility to ensure respect for human rights and to provide redress for violations of human rights.

This paragraph frames the obligations imposed by the treaty as obligations of result. Yet, if it is prevention of harm or risk of harm that is the result meant to be achieved, the obligations cannot be ones of result. Harm and risk of harm materialise all the time in relationship to various activities, and this harm and risk by themselves are never the *sole* basis for finding breach of obligations under human rights law.[105] The determination of breach is rather performed with reference to considerations of reasonableness, causation and State knowledge about the risk.[106] Reasonableness implies, for example, that even if the harm or the risk of harm was foreseeable (i.e., the State knew about it or ought to have known about it), and even if the State could have taken measures that causally could have prevented or mitigated the risk, the State might still not be in breach of its human rights law obligations. The reason is that it might not be reasonable to take measures,[107] since for example such measures might conflict with other legitimate and important interests.

If it is regulation of activities that is the result meant to be achieved, then indeed the obligation to regulate *as such* can be conceptualised as an obligation of result. That said, States still have discretion how to more specifically regulate, which makes the reference to a result (i.e., the result of regulation as such) not very helpful. Overall, then, not only does Article 4 of the CoE AI Convention fail to concretise, but the clarifications added to it by the Explanatory Report are confusing. As mentioned in Section II above, the Report is intended to provide some insights into the drafting process, without revealing the negotiating

---

[103] CoE Ad hoc Committee on Artificial Intelligence (CAHAI) Feasibility Study, CAHAI(2020)23, 17 December 2020, para 84.

[104] CoE Ad hoc Committee on Artificial Intelligence (CAHAI) Feasibility Study, CAHAI(2020)23, 17 December 2020, para 84. Concretisation would imply specifying that, for example, the right to fair trial includes a right to challenge evidence obtained with the help of an AI system.

[105] V Stoyanova, "Framing Positive Obligations under the European Convention on Human Rights Law: Mediating between the Abstract and the Concrete" (2023) 23(3) Human Rights Law Review 1. See also M Ambrus, "The European Court of Human Rights as Governor of Risk" M Ambrus, R Rayfuse and W Werner (eds) *Risk and Regulation of Uncertainty in International Law* (Oxford University Press, 2017) 100.

[106] See V Stoyanova, *Positive Obligations under the European Convention on Human Rights. Within and Beyond Boundaries* (Oxford University Press 2023); L Lavrysen, *Human Rights in a Positive State* (Intersentia 2016).

[107] States' positive obligations under human rights law to regulate activities by taking preventing or mitigating measures are not absolute in the sense of requiring the result of prevention and mitigation to be always achieved.

positions of the different States. Given the confusing formulation of the above quoted paragraph from the Explanatory Report, it can be inferred that the drafters were uncertain or could not agree how demanding and intrusive the obligations should be.[108]

The overall confusion as to the role of human rights law is also evident from the various drafts of the treaty. Here Article 6(2) of the Zero Draft deserves to be quoted in full. It stipulated that the State Parties shall

> ensure that the deployment and application of artificial intelligence systems in the public sector have an appropriate *legal basis* and that a careful *preliminary* consideration *of the necessity and proportionality* of the use of such system is carried out in view of the context of the deployment. (emphasis added).

This provision reflects the assumption that the use of AI systems in the public sector *inherently implies interferences with human rights*, which explains the imposition *per se* of the requirements for legal basis and for a necessity and proportionality assessment. It is the essence of human rights law to demand that any interference has a legal basis and that it is necessary and proportionate so that the interference does not constitute a violation.[109] In sum, the starting assumption in the Zero Draft was that the use of AI systems by public authorities has to comply with the standards of legality, necessity and proportionality, since the use *per se* interferes with human rights.

This starting assumption was diluted in the Revised Zero Draft to be eventually completely removed with the Consolidated Working Draft and in the final text of the treaty. To better understand the dilution, it is worthwhile to quote Article 5 of the Revised Zero Draft:

> Each Party shall, within its respective jurisdiction, ensure that:
>
> a. the application of an artificial intelligence system *substantially informing* decision-making by a public authority in the exercise of its function, or any private entity acting on its behalf, is fully compatible with its obligations to respect human rights and fundamental freedoms as guaranteed under its domestic law or under any relevant applicable international law;
>
> b. *any interference* with human rights and fundamental freedoms by a public authority or any private entity acting on its behalf *resulting from such application* of an artificial intelligence system is compatible with core values of democratic societies, in accordance with the law and necessary in a democratic society in pursuit of a legitimate public interest. (emphasis added).

The addition of "*substantially informing* decision-making" in Article 5 of the Revised Zero Draft adds two thresholds: a causation threshold and a severity threshold. Both are reminiscent of the difficulties and ambiguities surrounding the "high risk" classification in the EU AI Act, already addressed in more detail above.

---

[108] The distinction between obligations of means and obligations of result "rests on the extent to which international law encroaches upon the state machinery by instructing state organs to conform to a particular behaviour." A Ollino, *Due Diligence Obligations in International Law* (Cambridge University Press, 2022) 77; V Stoyanova, "Framing Positive Obligations under the European Convention on Human Rights Law: Mediating between the Abstract and the Concrete" 23(3) (2023) Human Rights Law Review 1.

[109] A conceptual distinction exists between the assessment that measures constitute an interference with important interests protected by human rights law and the assessment whether these measures constitute a violation of human rights law. See J Gerards and H Senden, "The structure of fundamental rights and the European Court of Human Rights" (2009) 7(4) International Journal of Constitutional Law 619.

The addition of these thresholds means that the assumption is not anymore that any usage of AI systems by public authorities inherently causes harm to important interests protected by human rights law, and it is therefore inherently an interference that has to have a legal basis and comply with the standards of necessity and proportionality. There is rather a causation threshold – the decision has to be informed by the system. Such a causation threshold is also reflected in the expression "resulting from such an application." Put it otherwise, the causal link between the usage of an AI system and any interference with human rights (i.e., harm to important interests) cannot be assumed. As to the severity threshold, it is reflected in the terms "substantially." In other words, there is a causality standard that needs to pass a certain threshold.[110]

The above quoted provisions were completely removed with the Consolidated Working Draft published on 7 July 2023,[111] a removal that, as suggested above, was related to the change and the limitation of the personal scope of the treaty (i.e., no need for specific provisions about public authorities since the Convention's scope was limited to regulating the conduct of States). The new provision on the scope of the treaty refers to AI systems that *have the potential to interfere* with human rights.[112] The accepted text of the treaty also uses the same expression. In particular, Article 3(1) of the CoE AI Convention provides that "[t]he scope of this Convention covers the activities within the lifecycle of artificial intelligence systems that have the potential to interfere with human rights, democracy and the rule of law [. . .]." The initial assumption that AI systems used by public authorities necessarily constitute an interference with human rights (and thus necessarily cause harm to important interests), was therefore replaced with the idea of potentiality to interfere. No causation or severity threshold are invoked in Article 3(1) of the Convention.

The resort to the causation and the severity threshold illustrates a general confusion and uncertainty about the role of human rights law. The Draft Framework Convention made public in December 2023,[113] whose text contains the various options available to the drafters to choose from, is very useful for showing this uncertainty. This Draft contained references to "violations of human rights,"[114] "adverse impacts on human rights,"[115] "unlawful harm or damage to the rights of individuals and legal persons,"[116] "the potential to interfere with human rights," "significantly affecting human rights,"[117] and "impacting on human rights,"[118] All these references reveal the difficulties in articulating, first, the role of human rights, second, the causal links between AI systems used by public authorities and harm to important interests as protected by human rights law, and, third, the causal links between AI systems used by private actors and the role of the State protect these important interests.

---

[110] On the different standards of causation as also related to the use of AI systems, see Sandy Steel, "Legal Causation and AI" in E Lim and P Morgan (eds), *The Cambridge Handbook of Private Law and Artificial Intelligence* (2024 Cambridge University Press) 189.

[111] CAI(2023)18, 7 July 2023.

[112] Art. 4, Consolidated Working Draft CAI(2023)18, 7 July 2023; Art. 3, Draft Framework Convention CAI(2023) 28, 18 December 2023.

[113] Draft Framework Convention, CAI(2023)28, 18 December 2023.

[114] Arts. 8, 14 Draft Framework Convention, CAI(2023)28, 18 December 2023.

[115] Ibid.

[116] Art. 14, Draft Framework Convention, CAI(2023)28, 18 December 2023.

[117] Ibid.

[118] Art. 15(1), Draft Framework Convention, CAI(2023)28, 18 December 2023.

3. Principles related specifically to the activities of AI systems

The second "legal gap" identified by the Feasibility Study is related the first one since it is also about specification. In particular, the second "legal gap" is intended to be addressed by means of explicitly legally enshrining certain principles that are specifically relevant to the operation of AI systems. These include human control and oversight, technical robustness, transparency and explainability.[119] Traceability via record-keeping and documentation of logs was also added,[120] which can enable the identification of any causal links between harms and the operation of the AI systems.[121] This in turn can enable the possibility of challenging the operation of the AI system under human rights law and the assessment whether there is any human rights law violation. Another principle that the Feasibility Study indicates as being specifically relevant to the operation of AI systems is ensuing that developers and deployers of AI systems have the necessary competence. The Feasibility Study observes that if these principles are not specifically regulated, this leads to uncertainty for both developers and deploys, which might hamper innovation.

The explicit legal enshrinement of certain principles that are specifically relevant to the operation of AI systems is intended to be achieved by Chapter III of the treaty. This Chapter enumerates principles "related to activities within the lifecycle of artificial intelligence systems." Given their nature as principles, they are "purposefully drafted at a high level of generality" and are meant to have "very broad application to a diverse range of circumstances."[122] As the Explanatory Report of the CoE AI Convention suggests, the drafters of the treaty assumed that the "detailed legal regime of human rights protection with its own set of rules"[123] is sufficient to further tailor the application of these general principles.

Yet, it is doubtful whether the current human rights regime can provide much specificity; it is still in the process of evolving to address the new challenges and harms to important interests arising from the use of AI systems.[124] Any obligations as corresponding to human rights are yet to be concretised. Such concretisation implies better understanding of the causal links between AI systems used by public authorities and harm to important interests as protected by human rights law, and, better understanding of the causal links between AI systems used by private actors and the role of the State protect these important interests. Such concretisation also implies proposals as to the measures (preventive or remedial) that might form the content of the human rights obligations. In the development of such proposals, the principles originating from the regulatory framework established specifically for AI systems is an important source of

---

[119] CoE Ad hoc Committee on Artificial Intelligence (CAHAI) Feasibility Study, CAHAI(2020)23, 17 December 2020, para 85.

[120] CoE Ad hoc Committee on Artificial Intelligence (CAHAI) Feasibility Study, CAHAI(2020)23, 17 December 2020, para 85.

[121] On the difficulties in establishing such causal links, see Sandy Steel, "Legal Causation and AI" in E Lim and P Morgan (eds), *The Cambridge Handbook of Private Law and Artificial Intelligence* (2024 Cambridge University Press) 189.

[122] Explanatory Report, para 49.

[123] Ibid, para 51.

[124] At the time of writing, the European Court of Human Rights has not delivered a judgment concerning harm that might have been caused by the usage of AI-systems. The closest areas covered so far in the Court's case law related to new technologies, concern mass surveillance and retention of data (see *Centrum för Rättvisa v Sweden* [GC] App no 35252/08 (ECHR, 25 May 2021) and facial recognition technologies (*Glukhin v Russia* App no 11519/20 (ECHR, 4 July 2023)). There is well-developed body of case law by the Court regarding data protection, which can be of relevance if the Court were to review usage of AI-systems (See Guide to the Case-Law of the European Court of Human Rights Data Protection (31 August 2024).). National courts, however, have made innovative interpretative advances. See M (Adamantia) Rachovitsa and N Johann "The Human Rights Implications of the Use of AI in the Digital Welfare State: Lessons Learned from the Dutch Syri Case" (2022) 22 Human Rights Law Review 10.

guidance. In this sense, it is the legal frameworks specifically regulating AI, such as the EU AI Act and the CoE AI Convention, that can help the development of human rights law.[125] Despite their high level of generality, the regulatory principles enumerated in Chapter III of the CoE AI Convention can be therefore helpful.

Being a source of guidance, these regulatory principles need to be better understood. This is the aim pursued in the following sections that consequently address the following principles: human dignity and individual autonomy, transparency and oversight, accountability, non-discrimination, data protection, reliability and finally, risk management. No comprehensive review is attempted. Rather some main contentious issues are identified in relationship to each principle.

a. Human dignity and individual autonomy

Article 7 of the CoE AI Convention enshrines the principle of respect for human dignity and individual autonomy. The Feasibility Study clarifies that "[t]o safeguard human dignity, it is essential that human beings are aware of the fact that they are interacting with an AI system and not misled in this regard." The Study adds that "they should in principle be able to choose not to interact with it, and to not be subject to a decision informed or made by an AI system whenever this can significantly impact their lives, especially when this can violate rights related to their human dignity."[126] As suggested in Section III.1. above, the Zero Drafts did contain specifically formulated rights to this effect.[127]

The Explanatory Report does not go that far. It clarifies among other things that respect to human dignity implies that individuals are not reduced to "mere data points."[128] According to the Explanatory Report therefore, individual autonomy does imply an ability to make choices and decisions and that individuals have "control over the use and impact of artificial intelligence technologies in their live."[129] The possibility to refuse to be subjected to a decision-making process by a public authority that might use an AI system, is however *not* mentioned as part of the conceptualisation of "individual autonomy" in the Explanatory Report to the treaty.[130] Nevertheless, human dignity and individual autonomy may provide important guidance for the contextual interpretation of human rights obligations. In certain contexts, the content of these obligations could be understood to include the right not to engage with an AI system or not to be subject to decisions generated by such systems.

---

[125] An important interpretative technique used by the European Court of Human Rights is resort to external legal frameworks, including EU law and CoE treaties, to determine the scope of the obligations under the ECHR. See Use of Council of Europe treaties in the case-law of the European Court of Human Rights (CoE, 2011).

[126] CoE Ad hoc Committee on Artificial Intelligence (CAHAI) Feasibility Study, CAHAI(2020)23, 17 December 2020, para 99.

[127] See also Article 9, Revised Zero Draft CAI(2023)01, 6 January 2023 that included the following clarification in its text as to human dignity: "in particular the ability to reach informed decisions free from undue influence, manipulation or detrimental effects which may adversely affect the right to freedom of expression and assembly, democratic participation and the exercise of other relevant human rights and fundamental freedoms as a result of the inappropriate application of an artificial intelligence system."

[128] Explanatory Report, para 53.

[129] Ibid, para 55.

[130] By way of comparison, it can be mentioned that the EU AI Act refers to dignity and individual autonomy in its preamble, but not in its actual provisions. It does oblige deployers under certain conditions to inform natural persons that "they are subject to the use of the high-risk AI system." See Article 26, EU AI Act. The possibility to refused to be subject to the use of a system is not, however mentioned.

b. Transparency and oversight

The second principle enshrined in the CoE AI Convention is transparency and oversight. In particular, Article 8 of the treaty stipulates that

> Each Party shall adopt or maintain measures to ensure that adequate transparency and oversight requirements tailored to the specific contexts and risks are in place in respect of activities within the lifecycle of artificial intelligence systems, including with regard to the identification of content generated by artificial intelligence systems.

A review of the different drafts related to transparency and oversight shows that there were no changes as to proposed formulations, which suggests that the final formulation of Article 8 of the treaty was not contentious. The only notable addition is the one made with the December 2023 Draft,[131] where identification of content generated by AI systems was added to the text. This was a positive addition, as also reflected in the actual text of Article 8 of the Convention, since it ensures that identification of such content is part of the transparency and oversight requirement. Although explainability is not mentioned in the text of the treaty,[132] the Explanatory Report suggest that explainability is part of the principle of transparency. The Report clarifies that explainability

> [.] refers to the capacity to provide, subject to technical feasibility and taking into account the generally acknowledged state of the art, sufficiently understandable explanations about why an artificial intelligence system provides information, produces predictions, content, recommendations or decisions, which is particularly crucial in *sensitive domains* such as healthcare, finance, immigration, border services and criminal justice, where understanding the reasoning behind decisions produced or assisted by an artificial intelligence system is essential. In such cases transparency could, for instance, take the form of a list of factors which the artificial intelligence system takes into consideration when informing or making a decision.[133]

As noted above, the Explanatory Report lacks binding authority; nonetheless, the above paragraph presents meaningful clarifications that offer interpretive value. As the text of Article 8 and the Explanatory Report suggest, certain "sensitive domains" can be identified, where transparency and oversight measures have to be tailored. This can imply more robust requirements. As clarified in the Explanatory Report, this tailoring could involve, among other things, enhancing transparency regarding the inputs or criteria considered by the AI system.

---

[131] Art. 7, Draft Framework Convention CAI(2023)28, 18 December 2023.

[132] Yet, see para 108 from the CoE Ad hoc Committee on Artificial Intelligence (CAHAI) Feasibility Study, CAHAI(2020)23, 17 December 2020: "they [those affected by the decision] should receive an explanation of how decisions that impact them are reached. While an explanation as to why a system has generated a particular output is not always possible, in such a case, the system's auditability should be ensured. While business secrets and intellectual property rights must be respected, they must be balanced against other legitimate interests. Public authorities must be able to audit AI systems when there are sound indication of non-compliance to verify compliance with existing legislation. Technical burdens of transparency and explainability must not unreasonably restrict market opportunities, especially where risks to human rights, democracy and rule of law are less prominent. A risk-based approach should hence be taken, and an appropriate balance should be found to prevent or minimise the risk of entrenching the biggest market players and/or crowding out and, in so doing, decreasing innovative socially beneficial research and product development."

[133] Explanatory Report, para 60 (emphasis added).

c. Accountability/responsibility, remedies and oversight mechanisms

Transparency can enable the identification of harms and the identification of actors that might have caused this harm, which is turn is indispensable for ensuring responsibility.[134] The CoE AI Convention thus specifically addresses accountability and responsibility.[135] In particular, Article 9 of the treaty is generally framed in the following way:

> Each Party shall adopt or maintain measures to ensure accountability and responsibility for adverse impacts on human rights, democracy and the rule of law resulting from activities within the lifecycle of artificial intelligence systems.

This might demand the adoption of new legal frameworks at national level or adaptation of the existing judicial, administrative, civil or other national liability regimes.[136] Such responsibility regimes are inherently linked with the right to effective remedy enshrined in Article 14 of the CoE AI Convention. Similarly to Article 8 of the treaty (transparency and oversight), Article 14 requires measures for documenting relevant information regarding AI system.[137]

There are qualifiers, though: such documentation is required when the system has "the potential to *significantly* affect human rights."[138] One can wonder what "significantly" means in this context and who decides whether the harm is significant or not. The same problem therefore emerges as the one discussed in Section IV above in relationship to Article 6(3) of the EU AI Act that excludes systems from the high-risk classification where they do not "pose a *significant* risk of harm to the health, safety or fundamental rights of natural persons."

One difference in articulation is the reference to risk in the AI Act provision. As observed in Section IV, the inclusion of risk weakens the required causal link between the system and the harm. As the available drafts of the CoE AI Convention reveal, challenges emerged in the drafting process how to articulate the causal links between AI systems and harm (or potential harm) to important interests protected by human rights law.[139] These challenges inevitably affect the possibilities for remedies and establishment of responsibility.

Addition qualifier comes with Article 14(2)(b) of the CoE AI Convention: States have to adopt measures to ensure that the information about the AI system is "*sufficient* for the affected persons to contest the decision(s) made or *substantially* informed by the use of the system, *and where relevant and appropriate,* the use of the system itself." The following questions are left open: Who decides when the information is *sufficient*? and Who decides whether the decision is *substantially* informed by the use of the system? Finally, no clear and straightforward obligation is imposed to inform about the use of the system itself. The commentary on Article 14 offered by the Explanatory Report seems to justify all these qualifications in the following way:

> It is also important to recall that exceptions, limitations or derogations from such transparency obligations are possible in the interest of public order, security and

---

[134] Ibid, para 69.

[135] It is not clear why the text refers to both "accountability" and "responsibility." The first one is normally associated with "broader and vaguer mechanisms." See Samantha Besson, "Theorizing International Responsibility Law, an Introduction" in Samantha Besson (eds) *Theories of International Responsibility Law* (Cambridge University Press, 2022) 1, 8.

[136] Explanatory Report, para 66.

[137] Article 14(2), CoE AI Convention.

[138] Ibid.

[139] For details see the last paragraph in Section V.2.

other important public interests as provided for by applicable international human rights instruments and, where necessary, to meet these objectives.[140]

Article 15 of the treaty strengthens the right to effective remedies by adding procedural safeguards. In particular, this provision stipulates that "where an artificial intelligence system *significantly impacts upon the enjoyment of human rights,* effective procedural guarantees, safeguards and rights, in accordance with the applicable international and domestic law" shall be ensured. Similarly to what was mentioned above, a threshold of significance (i.e., "significantly impact, significantly affect") is imposed.

This leads to the creation of two tiers of human rights harms: (1) normal and (2) significant. Human rights law generally does not incorporate such tiers and the right to effective remedy applies irrespective of the significance of the harm. In particular, once the definitional threshold of a right is engaged (i.e., it is determined that the interests protected by the right are impacted/harmed/interfered with),[141] remedies with procedural guarantees need to be ensured.[142] Indeed, the significance of the harm in terms of severity might imply more robust procedural guarantees. This is relevant since, for example, the right to private life is interpreted very generously. It is also relevant since the remedial procedural obligations corresponding to, e.g., Article 8 of the ECHR (private life) can be less robust in comparison to those corresponding to Article 3 of the ECHR (torture, inhuman and degrading treatment), in light of the different severity thresholds.[143] Yet, the key point remains that remedies and procedural safeguards are required in cases of both minor interferences and significant interferences with the enjoyment of human rights. There might only be variations in the robustness of the procedural guarantees.

Article 15(2) of the CoE AI Convention has an additional weakness. Its texts says that "Each Party shall *seek to ensure* that, *as appropriate for the context*, persons interacting with artificial intelligence systems are notified that they are interacting with such systems rather than with a human." The requirement to "seek to ensure" is weaker than "to ensure." The robustness of the requirements for notification is further undermined by the qualification "as appropriate for the context." It opens the possibility for an argument that notification in some contexts is not appropriate. Without notification, the affected person cannot know that an AI system has been used, which can undermine the possibility of subjecting the system to any scrutiny including via legal channels for establishing responsibility.

Finally, any discussion of responsibility must take into account the extent to which effective oversight mechanisms are available at domestic level in each State Party.[144] More generally, such mechanisms can include, national data protection authorities, equality

---

[140] Explanatory Report, para 99.

[141] See, generally, J Gerards and E Brems, "Introduction" in E Brems and J Gerards (eds), *Shaping Rights in the ECHR. The Role of the European Court of Human Rights in Determining the Scope of Human Rights* (Cambridge University Press, 2013).

[142] See, generally, J Gerards and E Brems (eds), *Procedural Review in European Fundamental Rights Cases* (Cambridge University Press, 2017).

[143] E Brems, "Procedural Protection: an Examination of Procedural Safeguards Read into Substantive Convention Rights" in E Brems and J Gerards (eds), *Shaping Rights in the ECHR. The Role of the European Court of Human Rights in Determining the Scope of Human Rights* (Cambridge University Press, 2013) 137.

[144] Note should be also taken of the Conference of the Parties, which is composed of representatives of the Parties to the Convention and can be considered as the oversight mechanisms *at the CoE level*. Yet, Article 23 of the treaty does not characterise the Conference of the Parties as a body with oversight mandate. Article 23 rather stipulates *inter alia* that the Parties shall consult with view to make "specific recommendations concerning the interpretation and application" of the treaty. Making such recommendations arguably entails some level of oversight.

bodies, human rights institutions with a general mandate or bodies specifically designated to oversee the application of AI systems. National legal systems differ as to the role of these mechanisms, including their standing and possible involvement in legal proceedings for establishment of responsibility.[145] And indeed, the CoE AI Convention *does not link*, on the one hand, Article 26(1) that obliges the State Parties to "establish or designate one or more effective mechanism to oversee compliance with the obligations of this Convention" with, on the other, Articles 14 (remedies) and 15 (procedural safeguards).[146] Although there is a clear obligation to have such mechanisms, States have discretion to determine their role.

Their role was weakened in the process of drafting the treaty. The Zero Draft contained the following proposed provision: "Parties shall establish or designate national supervisory *authorities* tasked, in particular, with overseeing and *supervising* compliance . . ."[147] As Article 18(2) of the Zero Draft suggested, supervising would imply a procedure that might lead to banning of systems. The possibility for banning of systems as part of the national authorities' supervising procedures were removed with the Revised Zero Draft.[148] Further modifications were introduced with the Consolidated Working Draft, where the reference to "authorities" was removed. The more general framing of "oversight mechanisms" was adopted.[149] The Draft Framework Convention published on 18 December 2023 removed the reference to supervision.[150] In this way Article 26(1) of the treaty took its final shape.

Although the Parties have discretion as to the type of mechanisms and the cooperation between them (if there are multiple national mechanisms), Article 26(2) of the Convention imposes two important requirements: first, independence and impartiality of the mechanisms, and second, conferral of power, expertise and resources for effective fulfilment of their overseeing tasks. The Explanatory Report clarifies that the first requirement implies "a sufficient degree of distance from relevant actors within both executive and legislative branches."[151] No distance from private actors is mentioned, which could be problematic given their possible influence and role in the setting of standards and technical specifications.[152]

d. Non-discrimination

Article 10(1) of the CoE AI Convention imposes an obligation to adopt measures for ensuring non-discrimination "as provided under applicable international and domestic law." Article 10 should be read together with Article 17 of the treaty that stipulates that "[t]he implementation of the provisions of this Convention by the Parties shall be secured without discrimination on any ground, in accordance with their international human rights obligations." These non-discrimination provisions do not add further specificities. As the

---

[145] See E Erken, The procedural participation of non-governmental organisations and national human rights institutions within the European Convention system and beyond: An empirical study (Doctoral thesis Utrecht University).

[146] See the contrast with Article 15 (Supervisory authorities) of the Convention 108+, where it is stipulated the supervisory authorities "shall have *the powers to issue decisions* with respect to violations of the provisions of this Convention and may, in particular, impose administrative sanctions" and that "shall have the power to engage in legal proceedings."

[147] Art. 18, Zero Draft CAI(2022)07, 30 June 2022.

[148] Art. 29, Revised Zero Draft CAI(2023) 01, 6 January 2023.

[149] Art. 25, Consolidated Working Draft CAI(2023)18, 7 July 2023.

[150] Art. 26, Draft Framework Convention, CAI(2023)28, 18 December 2023.

[151] Explanatory report, para 142.

[152] See Section V.3.f below.

Explanatory Report clarifies, the Drafters' intention was to "refer specifically to the body of the existing human rights law."[153] The intention was not to create new human rights obligations.[154] Whether existing obligations sufficiently address discrimination by AI systems is a matter of ongoing debate. While limitations have been identified, possibilities for the development of non-discrimination law have been also acknowledged as available.[155] There is a growing body of research in this area.[156] Article 10 of the CoE AI Convention, though not detailed, can offer a foundation that could support the advancement of anti-discrimination norms in this domain.

An interesting feature of the CoE AI Convention is its implicit engagement with what may be conceptualised as structural discrimination, as reflected in Article 10(2). Initially, the Zero Drafts referred to "rights related to discriminated groups and people in vulnerable situations."[157] The Consolidated Working Draft did not contain references to "people in vulnerable situations." It rather included a new paragraph that "called upon [the Parties] to adopt special measures or policies aimed at eliminating inequalities and achieving fair, just and equal outcomes," in line with domestic and international obligations.[158] The final version of Article 10(2) stipulates that "Each Party undertakes to adopt or maintain measures aimed at overcoming inequalities or achieve fair, just and equitable outcomes, in line with its applicable domestic and international human rights obligations." The Explanatory Report clarifies that with the treaty the State Parties aim at "overcoming structural and historical inequalities."[159] However, States preserve flexibility how to overcome these inequalities.

e. Data protection

Similarly to Article 10 (non-discrimination), Article 11 of the treaty that is about privacy and personal data protection, remains formulated at a relatively abstract level. This provision imposes two types of obligations. The first one is substantive – adoption or maintenance of measures to ensure that privacy and personal data are protected, "including through applicable domestic and international law, standards and frameworks." The second one is procedural – adoption or maintenance of measures to ensure effective guarantees "in accordance with applicable domestic and international legal obligations." The publicly available drafts suggest that the incorporation of these substantive and procedural obligations was widely accepted and remained uncontested during the negotiations.

While the Explanatory Report clarifies that States have discretion as to the substantive and the procedural measures,[160] it reveals that *the key role* of data protection laws was underscored during the drafting. Specific references are made to the CoE Convention on

---

[153] Explanatory report, para 71.

[154] Ibid, para 115.

[155] For a comprehensive study, see, e.g., J Gerards and R Xenidis, "Algorithmic discrimination in Europe: Challenges and opportunities for gender equality and non-discrimination law" (European Commission Report, 2020).

[156] See, e.g., S Wachter, B Mittelstadt and C Russell, "Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI" (2021) 41 Computer Law & Security Review 105567.

[157] Art. 5(3) of the Zero Draft, CAI(2022)07, 30 June 2022; Art. 12 of the Revised Zero Draft CAI(2023)01, 6 January 2023.

[158] Art. 9(2), Consolidated Working Draft CAI(2023) 18, 7 July 2023. Not changed in Draft Framework Convention CAI(2023)28, 18 December 2023.

[159] Explanatory report, para 77.

[160] Ibid, para 81: Art. 11 of the treaty "is not intended to endorse or require any particular regulatory measures in any given jurisdiction."

the protection of personal data (Convention 108+)[161] and the EU General Data Protection Regulation (GDPR).[162]

In the existing literature, AI systems have been reviewed from the perspective of data protection. Data protection law has imposed certain standards and requirements.[163] Such a perspective is very relevant since the development or the use of AI systems implies the processing of personal data.[164] The regulation of data protection is important with its detailed rules for lawful data processing and the rights of data subjects.[165] The GDPR sets out a comprehensive framework of principles regarding data protection, including the principles of transparency, purpose limitation and data minimisation to be applied when processing personal data.[166] Article 22(1) of the GDPR prohibits individual decisions from being taken in a fully automated manner.[167] One innovation brought with Convention 108+ is the conferral of a right upon individuals "not to be subject to a decision significantly affecting him or her based solely on an automated processing of data without having his or her views taken into consideration."[168]

Despite the importance of data protection laws, it has also been acknowledged that AI systems raise problems that transcend data protection.[169] In addition, data protection law has its own limitations.[170] The detailed nature of these limitations has attracted growing academic attention.[171] While data protection law has certain limitations, its focus on the rights of individuals whose data is processed, along with the evolving judicial developments,[172] ensures that it will continue to play a key role. As formulated, Article 11 of the CoE AI Convention, is intended to capture these developments through its reference to both domestic and international legal standards.

---

[161] Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (ETS No. 108) Strasbourg 28/01/1981, as amended with Protocol CETS No. 223 Strasbourg 10/10/2018. The aim of the Protocol of amendment is to modernise and improve the Convention (ETS No. 108), taking into account the new challenges to the protection of individuals with regard to the processing of personal data which have emerged since the Convention was adopted in 1980.

[162] Regulation 2016/679/EU of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data [2016] OJ L 119.

[163] Data protection regulations as a legal response have been useful for specifying how data can be collected, processed and stored. These regulations also provide rights to the data subjects. See Lena Enqvist, "Rule-based versus AI-driven Benefits Allocation: GDPR and AIA Act Implications and Challenges for Automation in Public Social Security Administration" (2024) 33(2) Information and Communications Technology Law 222.

[164] Recital (10), EU AI Act.

[165] McGregor, Murray and Ng, "International Human Rights Law as a Framework for Algorithmic Accountability" (2019) 68 International and Comparative Law Quarterly 314, 320, 325

[166] These principles can be important for the interpretation of Art. 8(2) ECHR. See, for example, the judgment by the District Court of the Hague, SyRI, 6 March 2020, para 6.41.

[167] There are many exceptions, however. Regarding Art. 22 GDPR see Case C-634/21 *OQ v Land Hessen* (Schufa) [2023].

[168] Art. 9(1)(a), Convention 108+.

[169] E Renieris, *Beyond Data: Reclaiming Human Rights at the Dawn of the Metaverse* (MIT Press 2023); F Palmiotto, "When Is a Decision Automated? A Taxonomy for a Fundamental Rights Analysis" (2024) German Law Journal 1.

[170] S Demkova, "The EU's Artificial Intelligence Laboratory and Fundamental Rights" in M Fink (ed), *Redressing Fundamental Rights Violations by the EU: The Promise of the 'Complete System of Remedies* (Cambridge University Press, 2025) 391, 415: "[...] data protection law guarantees data subjects' rights with substantial number of exception and limitations, which is evident from the long list of exceptions to the general prohibition on processing special categories of personal data in Article 9(2) GDPR. Such a priori exceptions might not be subject to the same proportionality and necessity test as permissible limits to fundamental rights are under Article 52(1) CFR."

[171] See, e.g., C Novelli, F Casolari, P Hacker, G Spedicato and L Floridi, "Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity" (2024) 55 Computer Law and Security Review.

[172] See, e.g., Case C-203/22 *Dun & Bradstreet Austria* [2025]. The ECtHR's case law has not yet developed in the area of data processing that includes AI systems. For a useful outline of ECtHR's cases relevant to data protection and the usage of new technologies, see Guide to the Case-Law – Data protection (31 August 2024) 95.

f. Reliability

Article 12 of the CoE AI Convention imposes an obligation upon States to take measures "as appropriate" to promote reliability.[173] These measures are meant to ensure "adequate quality and security" of the AI systems. The Explanatory Report suggests that these measures include adoption of standards, technical specifications, assurance techniques and compliance schemes. The Explanatory Report adds that Article 12 of the treaty is "based on the assumption that the Parties are best placed to make relevant regulatory choices."[174] This corresponds to the wide scope of discretion afforded to the State Parties under the treaty.

For ensuring reliability, there is a strong emphasis on standards and technical specifications. The advantage of standards is that they are concrete, in this way increasing clarity and certainty. At the same time standardisation as a method of regulation has been also criticised since it implies delegation of rule-making powers to non-representative or private bodies, which is not democratic and lacks legitimacy.[175] The Drafters of the CoE AI Convention have been sensitive to this problem. Paragraph 88 of the Explanatory Report reveals this sensitivity:

> In some cases, *it may not be enough to set out standards* and rules about the activities within the lifecycle of AI systems. Measures to promote reliability may therefore include, depending on the context, providing relevant stakeholders with clear and reliable *information* about whether artificial intelligence actors have been following those requirements in practice. This means ensuring, as appropriate, end-to-end accountability through *process transparency and documentation protocols.* There is a clear connection between this principle and the principle of transparency and oversight in Article 8 and the principle of accountability and responsibility in Article 9 (emphasis added).

While conformity with established standards may establish a presumption of regulatory compliance,[176] the obligation imposed by Article 12 of the CoE AI Convention of ensuring adequate quality and security requires more than standard-setting. It demands possibilities for questioning and revoking the compliance presumption. For these possibilities to be available there should be transparency and access to information. As noted in Section V.3.c. above, transparency can enable the identification of harms and the identification of actors that might have caused this harm, which is turn is indispensable for ensuring responsibility.[177] There is therefore an interconnection between the obligations for reliability and remedies.

g. Risk management

Article 16 of the CoE AI Convention stipulates that the State Parties shall adopt measures for the "identification, assessment, prevention and mitigation of risks posed by artificial intelligence systems by considering actual and potential impacts to human rights, democracy and the rule of law."[178] Such measures shall be "graduated and differentiated." States are therefore afforded flexibility in deciding which measures to implement, both in

---

[173] It is notable that the text of Art. 12 of the CoE AI Convention does not say that the State Parties take measures to *ensure* reliability, but only to promote reliability.

[174] Explanatory report, para 84–9.

[175] M C Gamito and C T Marsden, "Artificial intelligence co-regulation? The role of standards in the EU AI Act" (2024) 32 International Journal of Law and Information Technology 1.

[176] Ibid, 1, 7.

[177] Explanatory Report, para 69.

[178] Democracy and rule of law are not addressed here.

general terms and in relation to the specific domains where AI systems are used.[179] As clarified above, in contrast to the EU AI Act, the CoE AI Convention does not identify certain domains where the application of AI systems is considered high-risk.

Article 16(2) of the treaty elaborates on the required measures by specifying what they should *take into account* in light of the principles of graduation and differentiation. Specifically, the measures should consider "severity and probability of potential impacts" and "the perspective of relevant stakeholders." The provision also introduces *a temporal dimension*, stating that these measures should apply "iteratively throughout the activities within the lifecycle of the AI system." In addition, Article 16(2) clarifies *the nature of the required measures*: "monitoring for risks and adverse impacts to human rights, democracy, and the rule of law," "documentation of risks, actual and potential impacts, and the risk management approach," and "where appropriate, testing of AI systems before making them available for first use and when they are significantly modified."

Notably, the conceptual clarity of Article 16 is undermined by its inconsistent terminology. It refers variously to "risks to human rights," "adverse impacts to human rights," generic "risks," and "actual and potential impacts" – not all of which are explicitly tied to human rights. This inconsistency renders the article conceptually unclear. Further compounding the issue, Article 16(3) introduces the phrase "incompatible with the respect for human rights," without clarifying its relation to the terms used in the preceding paragraph in the same article.

Article 16(3) ensures discretion for the State Parties how to respond to the risks that might be identified. No obligation for banning system is imposed; rather, the State Parties adopt the procedural obligation to "assess the need for a moratorium or ban."

VI. Conclusion

Overall, the provisions of the CoE AI Convention reflect a wide scope of flexibility and high level of abstraction, which leaves discretion to its State Parties. This aligns with the design of the treaty as a Framework Convention that does not confer concrete rights upon individuals. This is a feature that distinguishes it from, for example, the CoE Convention Convention 108+ for the Protection of Individuals with Regard to the Processing of Personal Data. The choice of a Framework Convention was intended to *inter alia* ensure the participation of wide range of States and more willingness by States to become bound.

The high level of abstraction can be also linked to the choice of not incorporating concrete risk levels and tailoring regulations to the levels. In this respect, the Convention differs from the EU AI Act. Yet, given the ambiguities in the formulations of the risk levels and the possibilities for exclusions in the AI Act, the more general approach in the Convention does not necessary imply less robust regulation.

The regulatory impact of the Convention should be assessed in light of its interaction with human rights law, particularly given the specific mandate of the CoE – the institutional context in which the treaty was proposed and adopted. The Convention seeks to facilitate the effective implementation of human rights obligations,[180] which are still evolving in relation to the use of AI systems. Although the link between human rights law and the harm caused by such systems is not clearly conceptualised in the treaty's text, the principles set out in Chapter III offer interpretative guidance for the further development

---

[179] Explanatory report, para 106.

[180] It has to be also applied without prejudice to human rights law. See Art. 3(1)(b) of the CoE AI Convention: "When implementing the obligation under this subparagraph, a Party may not derogate from or limit the application of its international obligations undertaken to protect human rights, democracy and the rule of law."

of human rights norms in this area. For instance, the principles of human dignity and individual autonomy may play a central role in shaping the contextual interpretation of human rights obligations. Article 10 of the Convention, though not detailed, can serve as a foundation for advancing anti-discrimination norms in the context of AI. Similarly, Article 11 reflects ongoing developments in the field of data protection and may inform the articulation of legal claims involving the use of AI systems. Overall, although the review of the available negotiation documents indicates a reduction in the level of protection and in the scope and specificity of the obligations formulated in the treaty, the principles it articulates – human dignity and individual autonomy, transparency and oversight, accountability, non-discrimination, data protection, reliability, and risk management – can nonetheless provide interpretative guidance for the future development of human rights standards in the field of AI.