

THE DIFFICULT ROAD TO THE FRAMEWORK CONVENTION.
DILEMMAS AND STRENGTHS IN THE REGULATION OF
ARTIFICIAL INTELLIGENCE BY THE COUNCIL OF EUROPE

*Mario Hernández Ramos**

Abstract

The Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law, developed by the Council of Europe, is the world's first legally binding international treaty on this subject. Its main objective is to ensure that all activities within the lifecycle of AI systems are fully consistent with human rights, democracy and the rule of law. The Convention was designed to establish a universal legal framework. This was not only because non-member countries of the Council of Europe, including the United States of America, Canada, Japan and Israel, as well as Australia, Mexico, Costa Rica, Peru, Paraguay and the Holy See, played an important role in drafting and negotiating the Convention. It was also negotiated with the intention of sending a unified message to the world: the regulation of AI must be based on respect for human rights, democratic values and the principles of the rule of law, transcending particular approaches, interests and legal cultures. However, to achieve this result, the Committee on Artificial Intelligence, which negotiated and drafted the Convention, had to face a series of dilemmas that decisively shaped its content.

* Full Professor of Constitutional Law, Complutense University of Madrid; Chair of the Committee on Artificial Intelligence, Council of Europe.

TABLE OF CONTENTS

Introduction.....	1243
1. The Council of Europe before Artificial Intelligence.....	1244
1.1. The pioneering role of the Council of Europe as guardian of human values.....	1244
1.2. Ad hoc Committees on Artificial Intelligence: the CAHAI and the CAI.	1246
1.3. Brief introductory description of the Framework Convention.....	1247
1.4. The <i>status quo</i> of the Framework Convention: signatures and ratifications.....	1249
2. A Long and Difficult Road. The Dilemmas that Have Shaped the Framework Convention.....	1250
2.1. Introduction.....	1250
2.2. Is hard law regulation necessary?	1250
2.2.1. Regulation vs. Innovation: A false dichotomy	1251
2.2.2. Let`s ask the right questions: What to legislate, how to legislate and who should legislate?	1253
2.3 How detailed should the Framework Convention be legally?	1254
2.3.1. (New) human rights vs. principles.....	1254
2.3.2. The principles in the Framework Convention.....	1256
2.4. Should we be happy with the scope of the Framework Convention?	1261
Conclusion.....	1263

Introduction

Artificial intelligence is a technology that is currently present in every field, with a potential that has yet to be fully realised. Its use poses risks or directly impacts people's rights, which is why the recent technological boom has been accompanied by debate about its regulation. In this far from peaceful debate, advocates and detractors have clashed over its necessity, its appropriateness and the appropriate instruments. The Council of Europe, the largest international organisation focused on the protection of human rights, joined this debate and concluded that there is a need to regulate artificial intelligence in order to provide citizens with a legal instrument to deal with possible threats to their rights. The

main defining characteristics of this international intergovernmental cooperation organisation have made it possible to draw up a universal international standard in which 46 member and non-member countries (such as the United States of America, Canada, Japan, Australia, Israel, Mexico and Argentina, together with the European Union) agreed on the main regulatory standards for the use of this technology: the Framework Convention on Artificial Intelligence and human rights, democracy and the rule of law. However, the agreement between such diverse countries required negotiations that resulted in significant concessions.

The dilemmas faced during the negotiation and drafting of the Framework Convention shaped its content and design. Ultimately, in order to properly understand and assess the content of the first legally binding international treaty on this subject, it is essential to be aware of these dilemmas and the choices made by the parties. This is the main objective of these pages. Therefore, after briefly outlining the pioneering role that the Council of Europe has played in regulating artificial intelligence from the perspective of fundamental rights, democracy and the rule of law, the main dilemmas will be explained, through which the final content of the Framework Convention will be articulated.

1. The Council of Europe Before Artificial Intelligence

1.1. The pioneering role of the Council of Europe as guardian of human values

The Council of Europe was founded in 1949 as an international organisation with the main task of promoting democracy, human rights and the rule of law through law and the culture of cooperation between states on the European continent and beyond. It currently comprises 46 member states and around 700 million people, in addition to five observer states, namely the United States of America, Canada, Japan, Mexico and the Holy See. In pursuit of this mission, the European Convention on Human Rights and its guarantor body, the European Court of Human Rights, stand out among other internationally renowned treaties, such as the Personal Data Protection Convention and the Budapest Convention, in addition to an immense number of soft law instruments produced by its main bodies, such as the Parliamentary

Assembly and the Committee of Ministers, among others, and specialised committees, such as GRECO.

Without a doubt, the Council of Europe plays a prominent role in defending and promoting the essential nature of human beings as the supreme value that inspires all social organisations and their legal systems. This aspiration has been reinforced by technological developments, such as artificial intelligence, which are bringing about paradigm shifts across all areas in which the Council of Europe operates: human rights, democracy, and the rule of law. Consequently, it has taken a series of early initiatives to address this disruptive technology and be alert to both the risks and dangers it may pose, and the opportunities it may offer.

At its meeting on 17 May 2019 in Helsinki, the Committee of Ministers mandated its delegates to examine the feasibility and possible elements, on the basis of multi-stakeholder consultations, of a legal framework for the development, design and implementation of artificial intelligence, based on Council of Europe standards on human rights, democracy and the rule of law, thus establishing the Ad Hoc Committee on AI (CAHAI), whose first plenary meeting took place from 18 to 20 November 2019¹.

At its meeting on 21 May 2021 in Hamburg, the Committee of Ministers decided to prioritise work on an appropriate legal framework for the development, design and implementation of artificial intelligence, based on the Council of Europe's standards on human rights, democracy and the rule of law, and conducive to innovation. Finally, at its meeting on 20 May 2020, it welcomed the unanimous recommendation of the Committee on Artificial Intelligence that the Council of Europe introduce a legally binding cross-cutting instrument on artificial intelligence and human rights, democracy and the rule of law, in line with its mandate, "*which underlines the need for such an instrument to focus on general common principles, be conducive to innovation and open to participation by non-member states; [and] has instructed its Members to proceed rapidly with the elaboration of the appropriate instrument, taking into account other existing or developing international legal frameworks for artificial intelligence*".

Chapter drawn up within the framework of the research project "Fostering Artificial Intelligence for the reinforcement of the Rule of Law" with reference PID2024-157561NB-I00.

¹ <https://www.coe.int/en/web/artificial-intelligence/cahai>.

1.2. *Ad hoc* Committees on Artificial Intelligence: the CAHAI and the CAI

As mentioned above, the Committee of Ministers established two *ad hoc* committees on artificial intelligence: the Committee Ad Hoc on AI (hereinafter referred to as CAHAI)², whose mandate expired in December 2021 (from 11 September 2019 to 30 December 2021), and the Committee on AI (hereinafter referred to as CAI)³, which was constituted until the end of 2025 (from 1 January 2022 to 31 December 2025). As mentioned above, the main objective of these committees was to develop an appropriate legal framework to address the challenges that artificial intelligence may pose to human rights, democracy and the rule of law, along with a legally non-binding methodology for the risk and impact assessment of AI systems from the point of view of Human Rights, Democracy and the Rule of Law.

The CAHAI produced a “Feasibility Study” of elements of a legal framework for the design, development and implementation of AI based on Council of Europe standards. This was adopted on 17 December 2020⁴ and identified the main elements of a legal framework for regulating the design, development and implementation of AI, and a methodology for the risk and impact assessment of AI systems, both from the point of view of Human Rights, Democracy and Rule of Law.

This study formed the basis for the work of the CAI, which aimed to draw up a proposal for a horizontal legal instrument for the regulation of AI within the Council of Europe's competences. This objective was achieved at the tenth plenary meeting in May 2024. An interesting feature of the CAI is that it comprises not only the 46 member states of the Council of Europe, but also observer states: the United States, Mexico, Canada, the Holy See and Japan. In addition, other external countries also joined the CAI as observer states, including Australia, Israel, Argentina, Costa Rica, Peru, Uruguay and the European Union, represented by the European Commission. Additionally, 64 representatives of civil society, academia, industry and NGOs were part of the CAI, and contributed decisively to the CAI's debates and work. Thanks to this diverse composition, in terms of legal traditions and the

² <https://www.coe.int/en/web/artificial-intelligence/cahai>.

³ <https://www.coe.int/en/web/artificial-intelligence/cai>.

⁴ <https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-/1680a0c6da>.

particular interests of each party, the standards on AI, human rights, democracy and the rule of law enshrined in the Framework Convention, because of the dialogue and negotiation, enjoys significant legitimacy across much of the Western world. The Framework Convention is generating increasing interest from States, and since May 2024, the number of observer States has expanded, with Cameroon, Ghana, Ecuador, and Brazil now part of the CAI, and more states are expected to join⁵.

1.3. Brief introductory description of the Framework Convention

The Framework Convention consists of a Preamble and 36 articles, organised into eight chapters.

Chapter I establishes the “General Provisions” including the object and purpose (art. 1), the definition of AI systems (art. 2), and the scope of application (art. 3). Chapter II contains the “General Obligations” including the protection of human rights (art. 4), the integrity of democratic processes, and respect for the rule of law (art. 5). Chapter III includes the “Principles related to activities within the life cycle of artificial intelligence systems”, starting with a general orientation (art. 6), which includes human dignity and individual autonomy (art. 7), transparency and oversight (art. 8), accountability and responsibility (art. 9), equality and non-discrimination (art. 10), privacy and protection of personal data (art. 11), reliability (Art. 12), and safe innovation (art. 13). Chapter IV establishes “Remedies” (art. 14) and procedural guarantees (art. 15). Chapter V provides for a mechanism for “Assessment and mitigation of risks and adverse impacts”, detailing a minimum framework for risk and impact management (art. 16). Chapter VI provides guarantees in the “Application of the Convention” such as non-discrimination (art. 17), consideration of the rights of persons with disabilities and children (art. 18), public consultation (art. 19), literacy and digital skills (art. 20), and that the Convention will not entail a reduction in standards of protection of rights (art. 21 and 22). Chapter VII establishes a monitoring mechanism, called the *Conference of the Parties* (art. 23 and 24), and an obligation of international cooperation between the Parties (art. 25), as well as

⁵ It follows a scheme similar to the Council of Europe Convention 108, the first binding text on the protection of personal data, which was later adopted and developed by the EU and other States.

national oversight mechanisms (art. 26). Finally, Chapter VIII contains the final clauses for the termination of the international treaty, such as the possibility of making amendments to the Convention (art. 28), the manner of settling disputes between the Parties (art. 29), signature and entry into force (art. 30), accession after entry into force (art. 31), territorial application (art. 32), a federal clause for politically decentralised parties (art. 33), the possibility of making reservations only with respect to the federal clause (art. 34), denunciation (art. 35), and notification (art. 36).

To complete the summary description of the Convention, in addition to the content outlined above, we may identify five main characteristics.

First, a definition of AI is proposed that is functional, i.e. valid only in the terms of the Convention, and aims to be technologically neutral (an approach that inspires the entire Framework Convention), with the aim of ensuring that it remains valid despite new technological advances in the field of AI. Secondly, no new fundamental rights are created, but reference is made to existing rights at international and national level. Instead, the Convention consists largely of general principles that impose legal obligations on the parties, but at the same time allow and facilitate dialogue between European and non-European states with sometimes very different bills of rights. Thirdly, unlike the European Union's AI Regulation, it does not identify specific prohibited uses of AI but places the responsibility on the party to consider establishing moratoria, prohibitions or other measures when certain uses of AI systems are incompatible with respect for human rights, the functioning of democracy or the rule of law. Fourthly, a distinction is made between the public and private sectors with regard to scope. It will be up to the parties to decide how they comply with the obligation to 'address risks and impacts arising from activities within the lifecycle of artificial intelligence systems by private actors' (Art. 3.b). Lastly, States that are not members of the Council of Europe are permitted to ratify the Framework Convention, but if they have not been part of the negotiation process, they must wait for the Framework Convention to enter into force before they can ratify it.

1.4. The *status quo* of the Framework Convention: signatures and ratifications

Following the conclusion of the drafting and negotiation of the CAI, the Committee of Ministers of the Council of Europe adopted the Framework Convention on AI and Human Rights, Democracy and the Rule of Law on 17 May 2024 and opened it for signature on 5 September 2024.

At the ceremony held in Vilnius, ten States signed the Convention, including Council of Europe members such as Andorra, Georgia, Iceland, Norway, the Republic of Moldova, San Marino and the United Kingdom, observer States such as the United States of America, and States that participated very actively in the CAI, such as Israel. They were joined by the European Union, on its own behalf and on behalf of the 27 Member States. This means 36 States plus the European Union, which sent a clear and unambiguous message to the international community: it was possible to agree on essential and common principles for regulations to promote the responsible use of AI and that inspire, structure and guide national legal regulations.

Subsequently, and up to the date of writing, seven other countries have joined, furthering the universal profile that the Framework Convention aims to offer: Liechtenstein, Montenegro, Switzerland, Japan, Canada, Ukraine and Uruguay⁶.

Some states that did not participate in the Framework Convention's negotiation process have expressed interest in signing and even ratifying it. The intermediate step offered to them, which some have already adopted as previously mentioned, has been to become observer states within the CAI, with the intention of participating in its work and being in an optimal position for ratification when the Convention enters into force.

However, to date, neither the European Union nor any state has ratified the Framework Convention. Given that just five ratifications are required for the Convention to enter into force (art. 30 of the Framework Convention, and the fact that only Parties that have ratified the Convention are allowed to participate in drafting the rules of procedure for the Conference of the Parties, the Convention's monitoring body, it is expected that the first ratification will be followed by a wave of subsequent ratifications,

⁶ An updated list can be found at <https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence>.

making it easy to reach the threshold of five. Several signatory parties, including the European Union, have initiated their internal negotiation processes to proceed with ratification.

2. A Long and Difficult Road. The Dilemmas that Have Shaped the Framework Convention

2.1. Introduction

Although it has not yet entered into force, the existence of the Framework Convention is a milestone in today's political, legal and technological landscape. Given the circumstances, the fact that so many different countries agreed on, drafted and signed the first legally binding international treaty on AI and human rights, democracy, and the rule of law shows that, despite their differences, the members of the CAI wanted to send a loud and clear message of unanimity and unity on the approach and priorities for regulating artificial intelligence.

However, the final result offered by the Framework Convention is not without criticism and, like all human endeavours, is susceptible to improvement based on other approaches or attempts to achieve more specific interests.

Therefore, to understand the Framework Convention, it is essential to know how the path leading to this final result unfolded and the reasons for its content. To this end, it is necessary to explain that the Council of Europe, and the CAI in particular, faced several difficult dilemmas, which can be classified and systematised into three categories. The first dilemma concerns whether it was necessary or desirable to regulate artificial intelligence through a legally binding instrument. The second dilemma focused on the legal content of the Convention and how detailed it should be. The third and final dilemma concerns the scope and, in particular, how (through what legal mechanism) the commitments and principles of the Convention should be imposed on the private sector.

2.2. Is hard law regulation necessary?

The answer to this dilemma was already provided by the Feasibility Study prepared by the CAI and the position of the Committee of Ministers mentioned above, which favoured the drafting and negotiation of a legally binding international

instrument. However, this response coexists with a context of ongoing questioning about whether regulating this technology would hinder its development and, therefore, the realisation of its potential. In our opinion, the Council of Europe's position transcends the current context of the Framework Convention negotiations and reveals the importance of law and its values in a world where technology is ubiquitous and seems indispensable. Two key issues have been addressed in this dilemma: first, it must be stated that the dichotomy presented between innovation and regulation is false and self-serving, perpetuated by those who do not want regulation that subjects the exploitation, and therefore the self-serving use, of this technology to certain conditions. Secondly, once the false dichotomy has been denounced and overcome, and recognizing that regulation may raise problems, we must ask the right questions.

2.2.1. Regulation vs. Innovation: A false dichotomy

The regulation of artificial intelligence has generated heated debates since the beginning of the second decade of the 21st century. With the election of President Donald Trump to the White House, sharing positions with prominent business and technology leaders, there has been a resurgence of the view that regulating this technology represents a significant obstacle to its innovation and development. This perspective, while not new, was not the predominant view in the specialised forums where the regulation of artificial intelligence was debated in previous years. Nevertheless, most companies demanded a regulatory framework that would establish a level playing field in order to be able to invest and innovate with legal and business certainty.

But in the mainstream media, the false dichotomy prevails, and companies threaten to restrict their technological advances wherever regulations are unfavourable to them.

However, this stance ignores, perhaps deliberately, a key premise in the debate at hand. The starting point for addressing this dichotomy is undoubtedly the understanding that technology is not neutral, as is often claimed, despite mathematics and statistics being among its raw materials. Technology is a tool that pursues a specific goal, and that goal conditions and shapes the tool. With this premise in mind, and to summarise, artificial intelligence in our

societies already plays a political role⁷. Consequently, it cannot be left exclusively in the hands of companies that develop this technology according to their own agendas; rather, democratic bodies must play a leading role. These democratic bodies act in accordance with their constitutional frameworks, which mandate the protection and promotion of human values and democratic principles. This is because law is conceived as the most civilised way to resolve problems that may arise within a society. If regulation anticipates and resolves conflicts, what problem is there with regulating a technology that is already causing all sorts of problems and that is feared will be much worse in the future?

Therefore, the debate should not focus on whether artificial intelligence should be regulated or not. In fact, from a legal point of view, any legal system has sufficient tools to provide a response if the use of an AI system causes harm to a person. No judge in a similar dispute, claiming compensation for harm caused by the use of an AI system, can leave the dispute unaddressed. A different question is to what extent such an answer is satisfactory and how much judicial activism or legal creation the judge should develop. Judges, as a public authority, must make their decisions solely on the basis of the law. In this way, citizens can know in advance what arguments the arbiter of their dispute is likely to use and adapt their behaviour accordingly. In this way, arbitrary and abusive behaviour is also prevented. This requirement applies not only to judges, but to all public authorities in any democratic system, embodied in the legal principles that constitute the rule of law.

Without specific rules, people's behaviour and decisions will lack a legal frame of reference and the response of the authority will be unpredictable. Suitable and specific rules therefore provide certainty for all stakeholders, including companies, to protect their rights and interests and to counter arbitrariness, abuse of power and injustice.

In short, having a legal framework promotes respect for essential principles of the rule of law, such as the principles of legality and legal certainty that protect citizens from the abuse of those who hold a technological monopoly, whether private or state-owned.

⁷ For a more in-depth discussion of this idea, see M. Hernández Ramos (2025), *El impacto de la inteligencia artificial en el Derecho Constitucional*, Centro de Estudios Políticos y Constitucionales y Tribunal Constitucional, Madrid.

The debate should therefore focus on other types of questions that provide the rigour and complexity that this issue deserves, thus avoiding Manichean and simplistic postulates: Firstly, what should be the object of regulation, i.e. what should be regulated; secondly, how should it be regulated, i.e. by means of what instruments; thirdly, who should regulate it; and, fourthly and lastly, from what point of view, i.e. what ethical and/or legal principles should preside over such regulation, being aware that these principles will decisively condition its objective and its addressees.

2.2.2. Let's ask the right questions: What to legislate, how to legislate and who should legislate?

Answering these questions requires addressing new questions regarding the intended role of the Framework Convention. As will be seen, these questions are present in the three dilemmas discussed.

Firstly, it must be borne in mind that the Framework Convention will not be the only or the most important legally binding rule on the regulation of artificial intelligence. The Framework Convention is just one more rule in the complex and totally cross-cutting field of artificial intelligence, which is directly linked to the national and international regulations of the parties in the areas of human rights, democracy and the rule of law. The Convention, therefore, constitutes a normative element that complements possible gaps in other legal systems. This complementary relationship is evident with regard to the European Union's AI Regulation, given the obvious differences between the two regulations in terms of their legal nature, content, scope of application and purpose⁸. Secondly, and closely related to the above, the purpose of the Framework Convention must be taken into account, as it determines its content and the perspective from which the regulation will be designed.

⁸ On the relationship between these two standards, specifically on differences, similarities and their complementary relationship, see M Hernández Ramos (2024) El marco jurídico regulatorio europeo de la inteligencia artificial: La relación de complementariedad entre el Reglamento de la UE y la convención marco del Consejo de Europa. *Revista Española de Derecho Europeo* 92: 9-41, 17-36.

In accordance with the mandate of the Council of Europe, the Framework Convention must focus, as finally established in Article 1, on ‘ensuring that activities within the lifecycle of artificial intelligence systems are fully consistent with human rights, democracy and the rule of law’. This objective can be achieved in many ways; therefore, the margin of manoeuvre or margin of appreciation, so characteristic of the case law of the European Court of Human Rights, must be present in the drafting and content of the Convention. And this objective is best achieved through legal principles that impose obligations on States, rather than through the recognition of subjective fundamental rights that could hinder this necessary margin of manoeuvre. Lastly, the object of the regulation cannot be all uses of artificial intelligence, but only those that involve “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” (Art. 1.2). Consequently, legislation is enacted taking into account this approach to the risk and impact of artificial intelligence systems (Art. 16).

Thirdly and finally, consideration must be given to whether new fundamental rights or constitutional categories will be necessary for this new algorithmic reality. The answer is no, so far, but this will be addressed in the next point.

In short, the Framework Convention offers its own particular answer to the initial questions posed: first, it is necessary to regulate only certain use cases, those that pose a serious risk and impact on fundamental rights, democracy, and the rule of law. Second, regulation through an international treaty that establishes legal principles and allows for further national development enables greater scope for agreement among all parties of the CAI. And this flexibility is best utilised within a framework of intergovernmental cooperation, such as that offered by the Council of Europe, thus answering the third and final question.

2.3 How detailed should the Framework Convention be legally?

2.3.1. (New) human rights vs. principles

The content of the Convention is the second of the dilemmas that underpinned the work and negotiations in the CAI, not only

from a material but also from a formal point of view. This dilemma was resolved by posing a series of successive questions.

The first question to address this issue refers to the role that the Framework Convention would play in regulating AI. This question has already been developed in the previous point: in the negotiation and final drafting of the text, it was very important to consider that the Framework Convention would not be the only or the most important regulatory standard for artificial intelligence, but rather one more within the complex national and international regulatory landscape. The Framework Convention should be complementary to the already existing international human rights, democracy and rule of law standards and would aim at filling in any legal gaps that may have formed as a result of rapid technological advances in the sphere of human rights law but also with regard to the protection of democracy.

The second question concerns the objective and perspective of the Framework Convention's regulation. This question has also been addressed in the previous point: only those uses of AI systems that pose a significant risk and impact on human rights, democracy, and the rule of law should be subject to regulation.

To resolve this dilemma, two highly relevant and difficult questions had to be addressed, which are closely related. First, from the outset of the CAHAI's work, there was discussion about whether it was necessary to recognise new human rights, or new legal categories or principles, or whether the existing ones were sufficient. The discussion was continually fuelled by the different legal and political positions and particular economic, cultural and institutional interests of the CAI members: 64 representatives of civil society, industry and academia; 46 member States of the Council of Europe; 11 states with observer status in the CAI: United States of America, Canada, Mexico, Argentina, Costa Rica, Peru, Uruguay, Israel, Japan, Australia, Holy See; and the European Union. This heterogeneity gave rise to the second question: would it be possible to reach a final agreement on an instrument that imposed legal obligations on such different parties, with different legal traditions and from different geographical areas with their inherent cultural and political peculiarities?

Paradoxically, the problem inherent in the second question contained its solution: allowing countries outside the Council of Europe to participate in the drafting of the Convention demonstrated its intended global nature. Therefore, if the presence

of countries with legal traditions different from those of Europe, such as the United States, Canada, Japan, Israel or Australia, as well as Latin American countries (whose peculiarities with respect to Europe are much less significant than those of the other countries mentioned) was desired, the majority of European countries (and the European Union) could not expect them to renounce their legal approaches. Legal institutions as well established in Europe as the horizontal effectiveness of rights (the so-called German *Drittwirkung* or positive obligations in the European Court of Human Rights) or the principle of proportionality had no place in their legal systems. It was therefore necessary to negotiate with the aim of reaching agreement on common ground and setting aside differences. The logical consequence was that the agreements did not go beyond general agreements, essential in content but lacking in detail. The commitments in these agreements, so general and lacking in detail, could only be formulated as principles, which could be developed at national level, but not through fundamental rights with a much higher level of regulatory density and detail. This need for negotiation and agreement, together with the realisation that existing bills of rights, both national and international, were sufficient to fulfil the main objective of the Convention, and that no widespread demand for the crystallisation of a new right or legal category had been found in the specialised literature, determined that the Convention would be formed mainly of general principles.

2.3.2. The principles in the Framework Convention

For explanatory purposes, the different principles that make up the Convention could be grouped into four categories.

A) Framework principles: Chapter II (General obligations)

For explanatory purposes, the different principles that make up the Convention could be grouped into four categories.

The first group of principles would be those that provide structure and constitutional unity to the Convention, which constitute a true framework and impose framework obligations, if you will pardon the redundancy. Firstly, there is the principle of protecting human rights, which imposes the obligation to “ensure that the activities within the lifecycle of AI systems are consistent with obligations to protect human rights, as enshrined in applicable international law and in its domestic law” (art. 5).

Secondly, there is the principle of guaranteeing “the integrity of democratic processes and respect for the rule of law’ to ensure that artificial intelligence systems are not used to undermine the integrity, independence and effectiveness of democratic institutions and processes, including the principle of the separation of powers, respect for judicial independence and access to justice, and to ‘seek to protect its democratic processes in the context of activities within the lifecycle of artificial intelligence systems, including individuals’ fair access to and participation in public debate, as well as their ability to freely form opinions.” (art. 6)

The principles of democracy and the rule of law are contained in the same article, given their inseparable link: democracy only works if it is regulated by rules that respect the principles of the rule of law, and the rule of law only enjoys legitimacy if it arises from a democratic process.

This general obligation refers to two specific types of aspects.

Firstly, aspects of institutional relevance, whereby each party to the Convention must adopt or maintain measures to ensure that AI systems are not used to undermine the integrity, independence and effectiveness of democratic institutions and processes, including the separation of powers, respect for judicial independence and access to justice.

Secondly, it refers to processes of democratic relevance, in the sense that each signatory to the Convention must protect its democratic processes in the context of activities that fall within the life cycle of AI systems, including individuals’ free access to public debate and their ability to form opinions freely.

This perhaps somewhat convoluted expression refers to a proactive obligation on the part of the parties to monitor and ensure that people’s information environment is not manipulated in a way that conditions individual citizens in their freedom to vote, fighting against the ‘cognitive hacking’ that new technologies allow and facilitate, perverting the very essence of democratic participation.

In fulfilling their obligations to protect democratic institutions and processes, the Parties must focus on the risks posed by artificial intelligence systems to the principle of separation of powers between the executive, legislative and judicial branches; or to an effective system of checks and balances between the three branches of government; or else to a balanced distribution of powers between different levels of government (vertical separation of powers); or to political pluralism (with particular attention to

freedom of expression, freedom of association and freedom of peaceful assembly) and the existence of pluralistic and independent media and a range of political parties representing different interests and viewpoints, as well as fair access to and participation in public debate; otherwise to participation in democratic processes through free and fair elections, and plurality of forms of participation, together with respect for the rights of political minorities; or for respect for the rule of law (which generally encompasses the principles of legality, legal certainty and non-arbitrariness) and the principle of access to justice and its proper administration; and, of course, for respect for the principle of judicial independence.

B) Constitutional principles: Chapter III (Principles related to activities within the lifecycle of artificial intelligence systems)

These principles focus on protecting the most essential aspects of an artificial intelligence system throughout its life cycle. These are the principles of human dignity and individual autonomy (Art. 7), equality and non-discrimination (Art. 10), and protection of privacy and personal data (Art. 11). All of them are to be implemented “in regard to artificial intelligence systems in a manner appropriate to its domestic legal system and the other obligations of this Convention”.

C) Instrumental or technical principles

In addition to the aforementioned framework and constitutional principles, the Convention establishes a series of more technical principles that enable the effectiveness of the constitutional principles stated above. These are also included in Chapter III and are referred to as “Principles related to activities within the lifecycle of artificial intelligence systems.”

The first principle that should be mentioned due to its relevance is the principle of transparency, contained in Article 8 of the Framework Convention. This principle encompasses a series of technical and legal obligations of the utmost importance. As noted in the Explanatory Report to the Framework Convention, transparency refers to openness and clarity in the governance of activities within the life cycle of artificial intelligence systems and means that decision-making processes and the overall functioning of artificial intelligence systems must be understandable and accessible to the appropriate artificial intelligence actors. In some cases, this could also refer to providing additional information, for example, on the algorithms used, so providing transparency on an

artificial intelligence system could require communicating adequate information about the system. This implies, and is explicitly stated in Article 8 of the Framework Convention, identifying AI-generated content to avoid the risk of deception and to enable the distinction between authentic, human-generated content and AI-generated content, as the latter is becoming increasingly difficult for people to identify, through techniques such as labelling or watermarking.

Within the requirement for transparency, two key concepts must be highlighted: explainability and interpretability.

The term “explainability” refers to the ability to provide, depending on technical feasibility and taking into account the state of the art, sufficiently understandable explanations as to why an artificial intelligence system provides information, produces predictions, content, recommendations or decisions, which is particularly crucial in sensitive areas such as healthcare, finance, immigration, border services and criminal justice, where it is essential to understand the reasoning behind decisions produced or assisted by an artificial intelligence system. In such cases, transparency could, for example, take the form of a list of factors that the artificial intelligence system takes into account when reporting or making a decision⁹.

Another principle also contained in Article 8 of the Framework Convention is the principle of oversight, which refers to the various mechanisms, processes and frameworks designed to monitor, evaluate, and guide activities within the life cycle of artificial intelligence systems. These may potentially consist of legal, political and regulatory frameworks, recommendations, ethical guidelines, codes of practice, audit and certification programmes, and bias detection and mitigation tools.

Another important principle that must be upheld by parliamentary institutions is that of accountability and responsibility, as set out in Article 9 of the Framework Convention. These principles are inseparable from the principles of transparency and oversight mentioned above and refer to the need to provide mechanisms for individuals, organisations or entities responsible for activities within the life cycle of artificial intelligence systems to be accountable for adverse impacts on human rights, democracy or the rule of law resulting from activities within the life

⁹ Framework Convention Explanatory Report, p. 61.

cycle of such systems. Specifically, the provision requires Parties to establish new frameworks and mechanisms, or to maintain existing frameworks and mechanisms that may apply to activities within the life cycle of artificial intelligence systems to give effect to this requirement, which may include judicial and administrative measures, civil, criminal, and other liability regimes and, in the public sector, administrative and other procedures for challenging decisions, or the attribution of specific responsibilities and obligations to operators¹⁰.

Along with the more legal principles of equality and prohibition of discrimination (Art. 10) and privacy and personal data protection (Art. 11), we must conclude by highlighting the essential principle of reliability (reliability) (Art. 12), whose main purpose is to generate justified confidence in the reliability of an artificial intelligence system if its development and use comply with these standards.

This provision highlights the role that standards, technical specifications, assurance techniques and conformity schemes can play in assessing and verifying the reliability of artificial intelligence systems and in transparently documenting and communicating the evidence of this process. This principle aims to highlight the importance of establishing measures that seek to ensure the reliability of artificial intelligence systems through measures that address key aspects of performance such as robustness, safety, security, accuracy and performance, as well as functional prerequisites such as data quality and accuracy, data integrity, data security and cybersecurity. Relevant standards, requirements, and assurance and compliance regimes may cover these elements as a prerequisite for successfully building justified public trust in artificial intelligence technologies¹¹.

D) Remedies and procedural safeguards (Chapter IV)

Finally, reference can be made to the set of procedural obligations that seek to “ensure the availability of accessible and effective remedies for violations of human rights resulting from the activities within the lifecycle of an artificial intelligence system”.

This chapter is the most detailed in terms of specific actions required of the parties, which are specified in the following obligations: first of all, the obligation to document the relevant

¹⁰ Framework Convention Explanatory Report, p. 66.

¹¹ Framework Convention Explanatory Report, p. 84 and 85.

information regarding the AI system and its usage and to make it available to affected persons; second, such information should be sufficient to contest the decision(s) made or substantially based on the use of the AI system and the use of the system itself; third, the effective possibility to lodge a complaint with the competent authorities; fourth, the obligation to provide effective procedural guarantees, safeguards and rights to affected persons in connection with the application of an artificial intelligence system where an artificial intelligence system significantly impacts upon the enjoyment of human rights and fundamental freedoms; lastly, the provision of notice that one is interacting with an artificial intelligence system.

2.4. Should we be happy with the scope of the Framework Convention?

Without a doubt, the most difficult issue to agree on was the scope. Within the scope issues, aspects of national security (Art. 3.2) and the exclusion of research (Art. 3.3) were the subject of heated debate and tough negotiations. However, the issue that nearly derailed the negotiations and the agreement on the Convention itself was the inclusion of the private sector within the scope, and specifically, through which legal instrument the principles of the Convention (and which principles?) would be imposed on the private sector.

The difficulty and acrimony between the different parties was not only due to the political will to subject the private sector to certain legal regulations, but also to the difficulty of finding a legal formula that would fit into all legal and institutional systems, as mentioned above.

The final dilemma facing the CAI was whether to approve a Convention that allowed for differences in compliance with the principles contained therein between the public and private sectors, or not to approve a Convention, since the possibility of adopting the Convention by vote was not contemplated, but rather adoption by consensus.

The final agreement was reached in the evening on the last day of the final CAI plenary session, on 14 March 2024, after this issue, together with that of national security, had been discussed by small working groups that met regularly during the final plenary session.

The Framework Convention applies to both the public and private sectors. It allows the parties to do things differently, but the choice is theirs. Therefore, the Convention does not prohibit a Party from applying the same obligations to both the public and private sectors.

With regard to the public sector, the Framework Convention stipulates that 'each party must apply this convention to activities within the lifecycle of artificial intelligence systems undertaken by public authorities or private actors acting on their behalf' (Art. 3.1.a). However, with regard to the private sector, the Convention only obliges the parties to "address risks and impacts arising from activities within the lifecycle of artificial intelligence systems by private actors [...] in a manner conforming with the object and purpose of this Convention". (art. 3.1.b)

This obligation must be fulfilled taking into account three issues. First, this obligation must be consistent with the object and purpose of the Convention. In this regard, Article 1.2 establishes that each Party has to adopt or maintain appropriate legislative, administrative or other measures to give effect to the obligations set forth in this Convention. These measures must be ranked and differentiated as necessary in view of the severity and likelihood of adverse impacts on human rights, democracy and the rule of law during the life cycle of AI systems and may include horizontal measures applicable regardless of the type of technology used. Secondly, this obligation means that, upon ratifying the Convention, each Party must specify in a declaration how it intends to implement this obligation, either by applying the principles and obligations set out in Chapters II to VI of the Convention to private actors, or by taking other appropriate measures to comply with the obligations set out in Article 3.1.b). This declaration may be modified at any time or in any manner. Thirdly, each Party must report to the Conference of the Parties (the Convention's monitoring mechanism, developed in Chapter VII) within the first two years and thereafter periodically, on the details of the activities carried out to give effect to Article 3.1.a) and b) (Article 24).

This was the final outcome of the latest dilemma, which was heavily criticised by some sectors of civil society and the media, who claimed that it distorted the nature of a convention designed to protect human rights.

From the point of view of the Convention's complementarity with other legal instruments that can better regulate the private

sector, such as the EU AI Regulation (which was being debated and approved at that very moment in the European Parliament across the River Ill), the solution is very positive, in my view. It is undoubtedly much better than failing in the attempt to regulate artificial intelligence by the main Western liberal democracies and resigning ourselves to fragmented national legislation to deal with a cross-cutting and cross-border technological phenomenon.

Conclusion

Adopting the Framework Convention required all parties to negotiate and compromise on their demands and claims in order to reach an agreement. Above different legal traditions, perspectives and national interests, one idea led the negotiations and drafting: the need to give a united and universal political message that transcends national borders and interests, and to agree on common standards and principles for regulating AI. This message is that human values must underpin the design, development and deployment of artificial intelligence, and the Council of Europe Framework Convention plays a key role in achieving this. This is what the Convention truly embodies: agreement, understanding, and human values in the face of any threat.