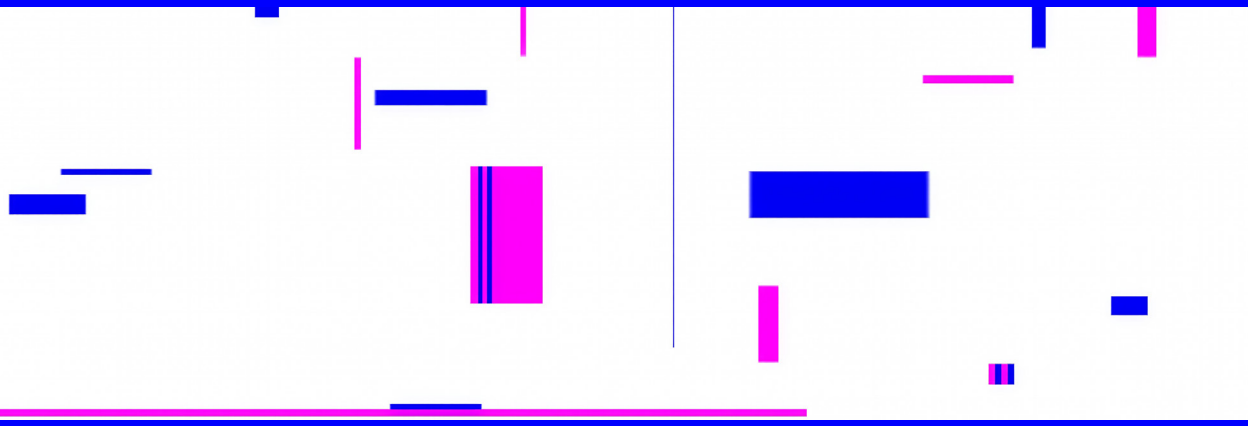
Incluye
Papel + Digital

Algorithmic Governance in Digital Health



Paula Subías Beltrán

Foreword by Itziar de Lecuona

www.bioeticayderecho.ub.edu

Universitat de Barcelona

unesco Càtedra de Bioètica de la Universitat de Barcelona

Observatori de Bioètica i Dret Universitat de Barcelona

ARANZADI

ARANZADI LA LEY, S.A.U.
C/ Collado Mediano, 9
28231 Las Rozas (Madrid)
www.aranzadilaley.es
Atención al cliente: https://areacliente.aranzadilaley.es/publicaciones

First edition: November 2025

Legal Deposit: M-24389-2025
ISBN print version: 978-84-1085-425-3
ISBN electronic version: 978-84-1085-426-0

Cuestiones bioéticas no resueltas en la evaluación de la investigación e innovación en salud basada en inteligencia artificial, tecnologías genéticas y datos personales (BIOEVAINNOLAW)". IP: Itziar de Lecuona. 1/9/23 – 31/8/27. Proyecto PID2022-138615OB-I00 financiado por:

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES
Cofinanciado por la Unión Europea
AGENCIA ESTATAL DE INVESTIGACIÓN

Design, Prepress and Printing: ARANZADI LA LEY, S.A.U.
Printed in Spain

Contents

II. THE VALUE OF NORMS

Resulta claro que las viejas recetas ya no funcionan, y que ya hace un tiempo dejaron de funcionar. Tampoco funciona la receta de la austeridad, que busca quitar al gobierno de en medio y espera que los mercados lo hagan todo sin un contexto definido claramente y una dirección segura y estable. Necesitamos un replanteamiento serio, una intensa creación de consenso, negociaciones mundiales y un liderazgo decidido. Las tecnologías capaces de impulsar una edad de oro global sostenible están al alcance de la mano: para desplegarlas con éxito solo hacen falta una comprensión del momento histórico y la voluntad de tomar una decisión sociopolítica clara.

- Carlota Pérez, *Otro capitalismo tiene que ser posible.*

The second part of this work examines the normative context relevant to the European setting in order to identify the fundamental questions that constitute the basis for the aforementioned algorithmic governance framework. The focus is on the European context because it is the point of origin and also because of its commitment to universal values such as human rights and democracy. It is a perspective that seeks not only to protect European interests but also to actively contribute to building a fairer and more sustainable world. This part begins by examining the value of establishing such a normative framework, focusing on the main governmental bodies with the capacity to guarantee the safeguarding of core values. Following this, significant reference works are identified, either because they are legally binding or have been recognized for their substantial impact on the regulatory framework, serving as the backbone of the framework. Then, a critical examination of these references is conducted with the objective of identifying the fundamental building blocks that constitute the algorithmic governance framework.

04

The european landscape

4.1. THE ROLE OF NORMS

In everyday life, we frequently engage in activities that bring us into contact with norms. Taking the subway or purchasing theatre tickets are acts that have normative significance, and in this case, legal significance, even if we are unaware of it: we can expect that the subway will take us to a specific location or that we are allowed into the theatre to see the play we are interested in. These instances, like many others, have a common feature that gives them normative meaning: we can demand specific actions from others, and others can expect specific behaviours from us [*Latorre*, 2019].

The law specifies which bodies and institutions are accountable for determining whether general behaviours are desirable and settling any problems that arise. These guidelines serve two purposes: to discourage particular behaviours and to promote specific expectations. These laws' authority is determined by the legitimacy of the bodies that create them rather than their intrinsic quality. The law seeks to make adherence to legal norms beneficial to those who follow them, either through the promise of a reward or the imposition of a punishment that entails the use of force. Consequently, some individuals are motivated to adhere to legal norms for practical and pragmatic reasons [*Atienza*, 1985].

But what is law? According to Miguel Reale, the law is a set of mandatory rules that ensure social coexistence by establishing limits on the actions of individuals [*Reale*, 1979]. Law is inherently a social phenomenon, existing only within the context of society. As a social construct, law is also dynamic, manifesting in various forms and continuously evolving while retaining elements of its historical foundations. A closer examination of the multiple definitions of the word *law* reveals that the various nuances

correspond to three key components of legal life: normative, factual, and axiological. Reale asserts that there is always an underlying fact (such as demographic or technical conditions), a value that assigns a specific meaning to that fact (aiming to its pursuit or rejection), and a norm. These elements do not exist in isolation; rather, facts, values, and norms coexist in a concrete unit. These factors are interdependent as components of a process, resulting in the dynamic and dialectical interaction that characterizes the law. If the underlying fact or value changes, the norm will change as well.

Within civilized societies, numerous norms dictate people's behaviour across various domains, spanning from societal trends to professional codes of conduct and legal rules. Law is limited to requiring the external observance of its rules insofar as they are necessary for coexistence and imposes an ethical minimum without which social life would be impossible [*Casado*, 2008]. The law must dictate rules that are valid for all, regardless of their moral opinions. These normative frameworks serve as mechanisms of social control, as transgressions against any of these rules entail adverse consequences for the individual involved. These consequences can range from societal disapproval and criticism to expulsion from professional associations, and potentially even legal ramifications. Our understanding of the functions that norms serve is, in turn, influenced by our own idea of society. The so-called functionalist approach holds that society tends toward equilibrium and thus regards legislation as a tool for resolving and preventing potential disputes [*Casado*, 2007]. They see society as a collection of different aspects that coordinate and integrate with one another and tend to balance. On the contrary, the conflictual view holds that society is fundamentally conflictive [*Casado*, 2007]. Hence, law is a tool that emerges from conflict and serves to lessen, rationalize, and conceal it. In any society, there can be both conflict and consensus, and the system of social control has an impact on both general and deviant behaviour.

In any case, the law's principal function is integrative, which means it reduces possible sources of conflict. The degree to which a law is applied, or, as it is commonly known, its *social validity*, varies greatly. Numerous examples throughout history demonstrate the use of legislation as a driver of social change [*Latorre*, 2019]. As previously said, it plays a direct function in certain circumstances and an indirect role in others; nevertheless, the impact on change is determined by a variety of criteria. For example, changes

enacted by law are more likely to be effective if the subject matter is neutral rather than involving values explicitly.

Ultimately, a legal system serves as a venture to regulate and structure social life, yet it never attains complete realization. There exists a certain tension between what is and what ought to be, and this tension between norms and reality is an inherent aspect of both legal and social domains. The inherent tensions of regulatory approaches may be exemplified by the regulation of the application of biology and medicine on people, as set forth in the Convention on Human Rights and Biomedicine [*Council of Europe*, 1997], commonly known as the Oviedo Convention. Since the end of the 1990s, these technologies have accelerated their course exponentially, so that the new techno-scientific possibilities have surpassed the ethical, legal, and political discourse on which the regulation that governed them was based. Therefore, it is now essential to review its content, both in terms of the issues it addresses (some of which stem from past concerns that have since been resolved) and those it omitted (due to their unforeseeable nature or attempts to reconcile diverse viewpoints). In this context, it is important to note how social and legal-political sensitivities have evolved alongside techno-scientific progress. In some instances, society has accepted new scenarios and their accompanying questions, while in others, it has firmly and dogmatically opposed new technologies. Additionally, the greater the disparity between the current regulation and the social and scientific realities in which it is applied, the more challenging it will be to adapt the norms to current issues [*Casado and López Baroni*, 2021].

A common idea is that regulation constrains research, but preventing this framework from becoming restrictive depends on our openness to change and our ability to manage reporting and control mechanisms transparently through regulatory systems [*Casado*, 2011a]. The law introduces a factor of rationalization and certainty, and it serves a function of legitimization and control. However, neither international conventions nor national laws alone can provide direct answers to the questions raised by technoscientific progress [*Casado*, 2008]. Since society, public authorities, and the legal system must make decisions without fear and ignorance, it is essential to create new ethical bodies and forums for discussion involving relevant stakeholders. For democracy, this is an educational approach that ensures citizen participation by establishing spaces for reflection and action. In these circumstances, information and social debate become

essential prerequisites for regulatory work. The rigor and richness of this discussion will determine whether the adopted solutions align with the values deemed relevant by society and respect minority options. Indeed, there can be *valid reasons* for diverse viewpoints, and this lack of ethical-social consensus often leads to a demand for legislation, assigning the law the role of settling discussions on issues without a clear social consensus, thus reinforcing the demand for legislation. It is undeniable that the legal system has the function of addressing and resolving conflicts [*Bobbio*, 1980]. However, the existence of a law does not definitively settle an issue: social debate persists, and the application of the law can give rise to new conflicts. Therefore, those dealing with key bioethical issues in public law must understand not only what the law says but also the moral implications and public perception related to these issues, meaning they must rely on bioethical reflection.

4.2. REFERENCE WORKS

The ethics of science and technology has acquired a relevant role in defining European Union policy, both at the European Union level and within individual Member States. While the European Union must respect the competencies of its Member States, it must also conduct its own reflection and regulation, which is often difficult [*Salvi*, 2009]. This section analyses the normative instruments for AI technologies from a European perspective, with a particular focus on the Council of Europe and UNESCO.

The normative corpus that fosters AI governance in Europe has evolved substantially over the last few decades as ethical concerns about this technology have grown. The reporting of adverse events has played an important role in reflecting on our standards of conduct, emphasizing the need for a robust regulatory framework. These occurrences have not only shaped our ethical code but have also fuelled the call for more restrictions to ensure that AI development and deployment are consistent with society's values and public safety. As a result, European institutions are increasingly focused on developing norms to address the complexities and potential threats linked with AI. Fig. 4.1 shows a timeline with a selection of incidents related to algorithmic discrimination [*Landon et al.*, 2005; *Larson and Angwin*, 2016; *Dieterich et al.*, 2016; *Buolamwini and Gebru*, 2018; *Amnesty International*, 2021], misrecognition [*Lee*, 2009; *Carty*, 2011; *Barr*, 2015], neglected self-determination [*Yang et al.*, 2020], limits of

privacy [*Grabham*, 2013], absence of monitoring protocols [*Conger*, 2020; *Hunt*, 2016], concerns about the environmental impact of AI [*Anthony et al.*, 2020], and AI impersonating humans [*Ngamkham*, 2025; *Harrison*, 2023]. These situations have generated a growing awareness of the risks associated with the indiscriminate implementation of AI technologies and have driven the public demand for stricter and more ethical regulation in this field.

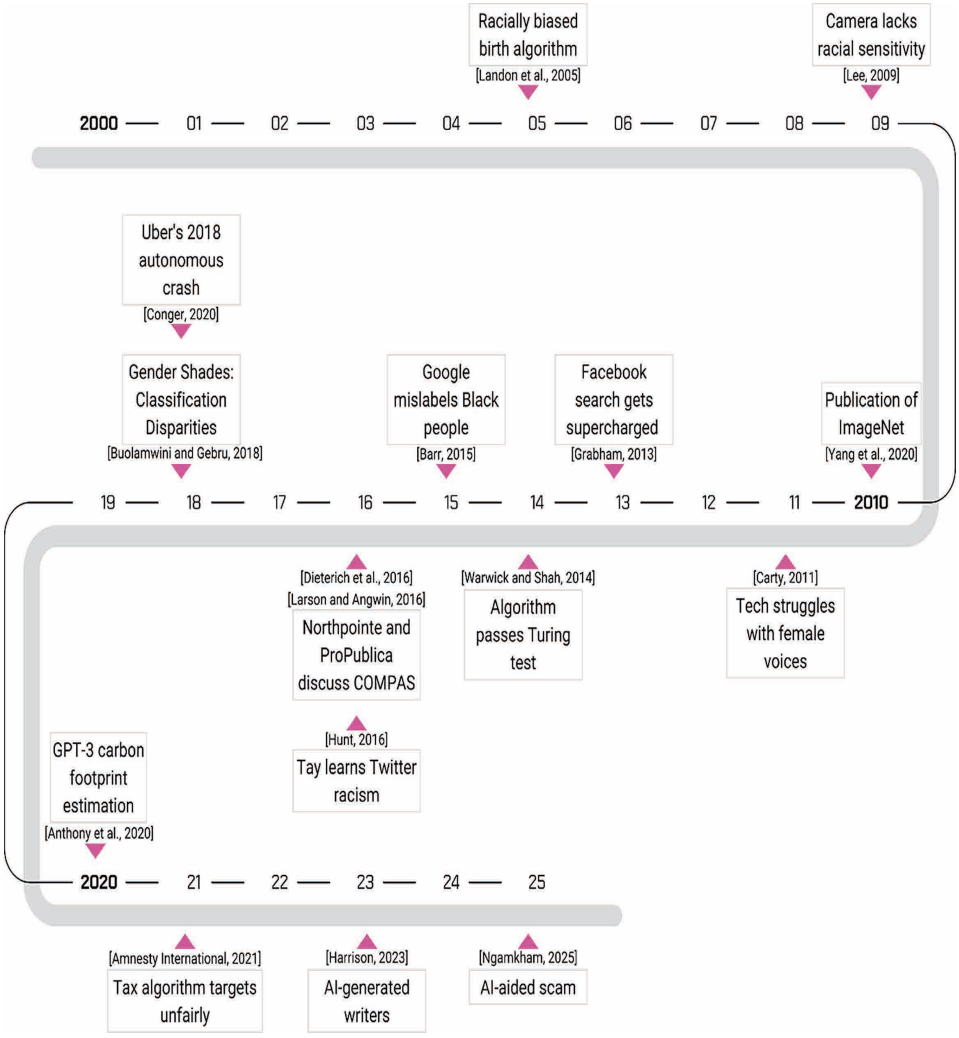


Fig. 4.1. Selection of adverse events that have occurred in the last years with AI-based solutions.

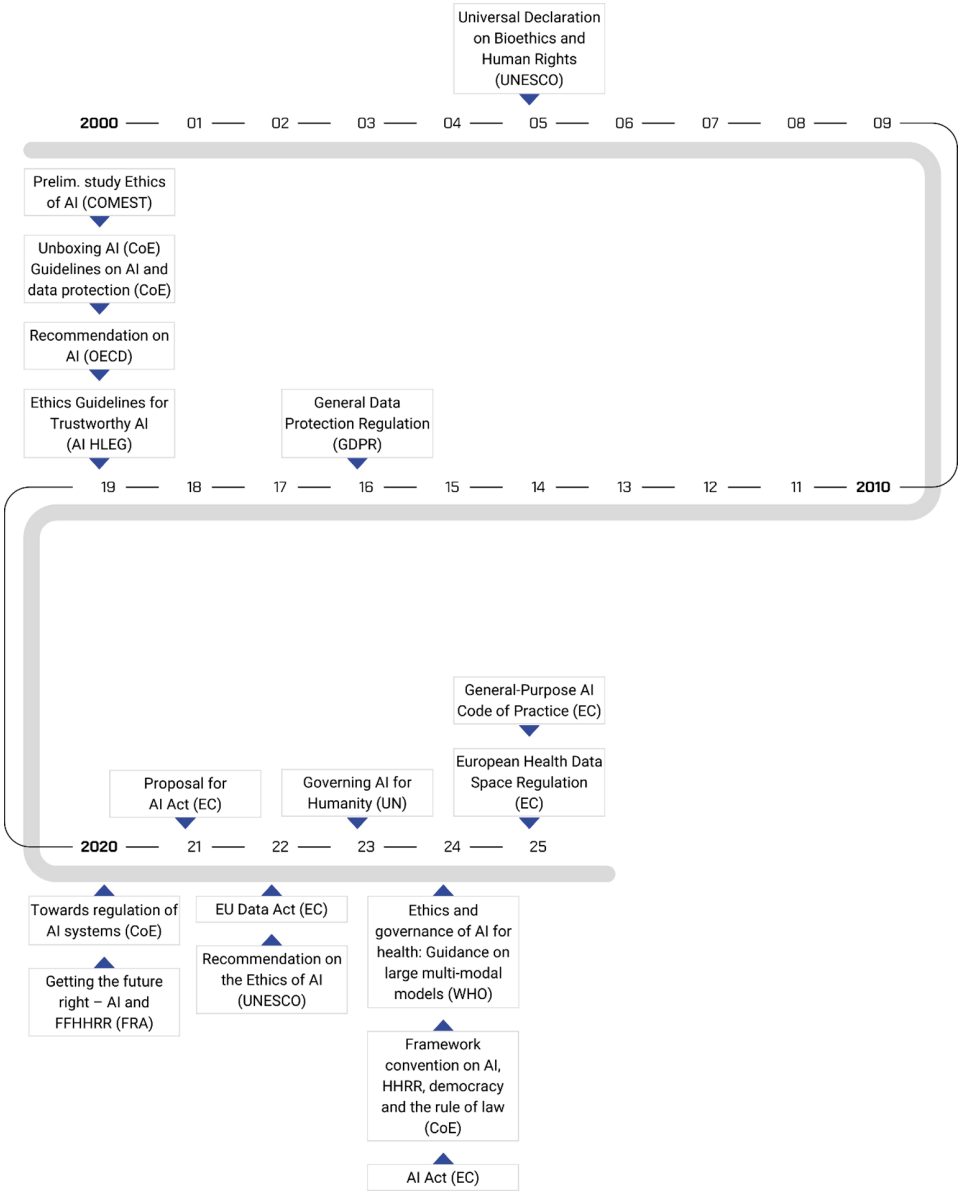


Fig. 4.2. A selection of reference works from governmental organizations influencing AI governance in Europe.

In response to these challenges, the normative instruments have evolved to more effectively address the complexities and risks inherent in AI. This evolving process reflects an ongoing effort by the international community to

adapt to technological advances and promote that AI is used ethically and for the benefit of society as a whole. The works of reference for this research are authored by governmental institutions that influence the European context, like the European Commission (EC), the Council of Europe (CoE), the United Nations (UN), the World Health Organization (WHO), or the Organization for Economic Cooperation and Development (OECD).

Fig. 4.2 depicts the principal works on AI that have been produced by the governmental institutions of interest. The selected texts are presented in chronological order and subsequently discussed in a similar manner. Whenever feasible, the works are grouped by authoring institution to facilitate comprehension. This section provides an overview of the motivation and principal contributions of each work. The analysis of these works is ultimately aimed at establishing a general context for the discussion at the normative level. Subsequently, *Chapter 5* will narrow the scope of the analysis to the specific contributions of the Council of Europe and UNESCO, based on the Council's binding legal framework and broad mandate, and UNESCO's strong emphasis on the protection of human rights. It is evident that there is some degree of overlap, which serves to illustrate the over normative ecosystem that currently characterizes AI.

The founding of the Council of Europe in 1949 meant the creation of an international entity to uphold human rights, democracy, and the rule of law in Europe. One of the best-known bodies that hangs on the Council of Europe is the European Court of Human Rights, an international court that interprets the European Convention on Human Rights. Another body of the Council of Europe is the Directorate General Human Rights and Rule of Law, which has overall responsibility for the development and implementation of the human rights and rule of law standards of the Council of Europe. And there is also the Commissioner for Human Rights, an independent and impartial non-judicial institution dedicated to engaging in dialogue with national authorities and civil society, and analysing, advising, and raising awareness of systematic human rights work. In January 2019 the Directorate General Human Rights and Rule of Law published a proposal of guidelines [*Directorate general of human rights and rule of law*, 2019] while in May 2019 the Commissioner for Human Rights issued a Recommendation [*Council of Europe - Commissioner for Human Rights*, 2019] that addressed the risk of ML-based systems undermining human rights rather than strengthening them. The approach of both proposals is to empower individuals by ensuring that they can understand the why of AI-powered decisions as well as their verification process. Besides, to increase reliability in AI, they encourage

AI developers and vendors to design products in such a way that they safeguard users' freedom of choice over the use of AI.

In 1998 UNESCO set up the World Commission on the Ethics of Scientific Knowledge and Technology (also known as COMEST), with the ambition of creating an advisory body for decision-makers, a body capable of whistleblowing about risky situations, and a forum of reflection. In 2019 COMEST published a preliminary study on the Ethics of AI [*COMEST*, 2019], which included multiple ethical considerations about AI development and suggested several elements that could be included in an eventual Recommendation on the Ethics of AI. In 2007 the European Union Agency for Fundamental Rights was established, which is better known in English as the Fundamental Rights Agency (FRA). The FRA is a European Union-independent body that helps to safeguard the rights, values, and freedoms that set the European Union's Charter of Fundamental Rights at the European Union, national, and local levels. Due to the role that AI plays in many decisions that affect our daily lives, in 2020 FRA published a report called "Getting the future right - Artificial intelligence and fundamental rights" [*EU Agency for Fundamental Rights*, 2020] where they provided an overview of the use of AI in the European Union, an analysis of the awareness of fundamental rights and further implications, and a discussion of measures to assess and mitigate the impact of AI on people's fundamental rights. Among others, this report increases the understanding of how AI-based solutions may cut across different rights.

In 2019 the *High-level expert group on AI* (AI HLEG), which was appointed by the EC for the 2018-2020 period, published the "Ethics Guidelines for Trustworthy AI" [*AI HLEG*, 2019]. This work, where they list seven key requirements that AI systems should meet in order to be *trustworthy* based on human rights and ethical principles, has been central to the development of the EC's approach to AI, although not without critique[1]. Shortly after, the OECD published its "Recommendation on AI" [*OECD*, 2019]. This document is focused on the promotion of a human-centric approach to trustworthy AI, while fostering research, preserving economic incentives to innovate, and assuring the representation of all stakeholders.

The *Ad hoc Committee on Artificial Intelligence* (CAHAI) was established in 2019 by the Council of Europe's main decision-making body and tasked

1. There was a debate about the voices represented in its creation, with a minority of four ethicists and significant industry representation [*Metzinger*, 2019]. Should it be the voice of industry the main one pulling the strings?

with examining the feasibility and possible elements of a legal framework for the development, design, and application of AI, based on Council of Europe's standards on human rights, democracy and the rule of law, which was published in 2020 under the title "Towards regulation of AI systems" [*CAHAI*, 2021]. This publication aimed to feed the ongoing reflection within CAHAI on the analysis of the challenges posed by AI systems and possible regulatory responses.

In 2022 the *United Nations Educational, Scientific and Cultural Organization* (UNESCO) published a seminal reference work, the "Recommendation on the Ethics of AI" [*UNESCO*, 2022]. This recommendation[2] focuses on the practical applications of AI, providing a framework for policymakers to translate the core values and principles into action with respect to a range of issues, including data governance, environmental and ecosystem management, gender equality, education and research, and health and social wellbeing. One year later, the United Nations published a report entitled "Governing AI for Humanity" [*UN's AI Advisory Board*, 2023]. In this report, the United Nations discusses the opportunities and challenges brought about by AI and outlines a strategy for AI governance. This strategy is designed to ensure that the potential of AI is harnessed in a way that leaves no one behind. The *Committee on Artificial Intelligence* (CAI), which was founded in 2022 by Council of Europe's Committee of Ministers to continue the work carried out by CAHAI, was tasked with examining legal frameworks to govern the creation, design, and implementation of AI systems, with the goal of drafting a Convention aligned with the Council of Europe's standards [*Committee on Artificial Intelligence*, 2023]. The proposed "Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law" was the result of two years of work that brought together the 46 Member States of the Council of Europe, the European Union, and 11 non-member states, as well as representatives of the private sector, civil society, and academia who participated as observers [*Council of Europe*, 2024]. In the words of the Council of Europe's secretary, the Convention represents "a response to the need for an international legal standard supported by states in different continents which share the same values to harness the benefits of AI, while mitigating the risks". The Framework Convention was adopted on May 17th, 2024.

2. Emanating from the Organization's supreme governing body and hence possessing great authority, recommendations are intended to influence the development of national laws and practices [*UNESCO*].
These are norms which are not subject to ratification, but which Member States are invited to apply.

Additionally, the WHO also made its own statement by proposing a specific governance framework for large multimodal models (LMMs) in healthcare [*WHO*, 2024]. Aimed at governments, developers and health professionals, the guidance provides over 40 recommendations to ensure that LMMs are used safely, equitably and effectively. It highlights five key areas of use while addressing significant risks such as misinformation, bias, data quality issues, cybersecurity threats, and automation bias. WHO emphasizes the need for transparent regulation, stakeholder engagement and rigorous oversight to align the development and use of LMMs with human rights and public health goals.

Complementing these global efforts, the European Union has taken a leading role in regulating AI through the AI Act, a landmark legislative framework that seeks to comprehensively govern the development, deployment, and use of AI across member states. With a risk-based approach, the Act sets out obligations proportionate to the potential harms posed by AI systems, including those used in healthcare. Its ambitious scope extends beyond the EU's borders, influencing global standards by requiring compliance from any AI system that affects individuals within the EU, regardless of where it is developed. The AI Act aims to ensure that AI technologies uphold fundamental rights, transparency, and safety throughout their lifecycle. The first draft was released in 2021 [*European Commission*, 2021] and has been revised several times since then [*Artificial Intelligence Act*, 2024].

Thus far, the analysis has focused on the normative documents that explicitly address AI-based technologies. However, it is relevant to recognize the influence of other regulations that indirectly shape the development of this technology due to its reliance on data, particularly in the context of ML. This includes the GDPR, the European Union Data Act, and the European Health Data Space. The GDPR establishes strict guidelines for the protection of personal data, directly affecting how information is collected, stored, and used, ensuring the privacy and security of users. On the other hand, the European Union Data Act seeks to facilitate the access and use of data in the European Union, promoting an environment in which information can be shared and reused securely and efficiently, which is crucial for the development and effective implementation of AI-based solutions. Conversely, the European Health Data Space (EHDS) is a flagship EU initiative aimed at enabling secure and interoperable access to health data across member states. It seeks to empower individuals with greater control over their health information while facilitating the responsible use

of data for research, innovation, and policymaking. By supporting cross-border data sharing and harmonized governance, the EHDS provides a critical foundation for the ethical development and deployment of digital health technologies, including AI systems. Thus, although these regulations do not focus exclusively on AI, their impact on the field is significant and cannot be ignored.

Particular attention is given to the analysis of the recommendation and Convention suggested by [*UNESCO*, 2022] and [*Council of Europe*, 2024], with UNESCO's Recommendation considered the most relevant standard due to its strong emphasis on the protection of human rights. However, it is just a symbolic and non-binding statement. The Council of Europe's Convention on AI comes next due to its scope and timely adequacy. Nonetheless, its binding nature will only apply to the nations who willingly sign on to it.

This work prioritizes the Convention proposed by the Council of Europe over the European Union's AI Act, primarily due to its broader human rights-oriented scope. While the AI Act represents a significant regulatory effort within the EU, its primary focus lies in risk management and the harmonization of market standards across Member States. In contrast, the Council of Europe's Convention is grounded in a comprehensive human rights framework, extending its relevance beyond market considerations to include wider ethical, legal, and societal dimensions. Its pan-European reach and emphasis on safeguarding fundamental rights make it a more suitable instrument for evaluating the implications of AI technologies, particularly in sensitive sectors such as health. Furthermore, the decision has been taken to prioritise anachronic texts, which provide a solid foundation for the proposed governance framework, ensuring a deeper and more enduring ethical basis.

By primarily focusing on these two publications, a comprehensive and reliable perspective will be obtained, integrating the most protective and recent thoughts and suggestions within the prevailing normative framework [*López Baroni*, 2024]. This approach provides a deeper understanding of AI's ethical and regulatory concerns. To facilitate the discussion, the broad topic will be addressed by focusing on specific concerns raised by the normative landscape of reference. While the substance of published documents may differ, the fundamental ideas can be identified and mapped, as shown in the next chapter.

05

Normative framework review

5.1. THE SCOPE

In the realm of advancing technology, a profound commitment to human rights forms the bedrock of UNESCO's approach [*UNESCO*, 2022]. They acknowledge the transformative power of AI, recognizing its potential to shape our world. Yet, in doing so, they firmly believe that addressing risks and ethical concerns should not stifle innovation but rather serve as a compass guiding us toward a future where progress aligns seamlessly with human values. Central to their philosophy is the recognition that a minimum body of ethical standards can act as a keystone in forging AI-related norms across the global landscape. Anchoring themselves in established frameworks, they provide a solid foundation for ethical discourse. This recommendation aims to be an international standard-setting instrument on the ethics of AI with a recommendatory character, acknowledging the jurisdiction of UNESCO. They argue that AI development should be directed not only by solid scientific research, but also by rigorous ethical analysis and review. And they intend to go beyond the identification of key values and principles by developing concrete policy proposals that allow readers to move from theory to practice. In the symphony of progress and ethics, their narrative harmonizes human rights with the boundless potential of AI, envisioning a future where innovation and morality walk hand in hand.

The primary task instructed to the Committee on AI appointed by the Council of Europe was to elaborate a suitable legal framework that complied with the Council of Europe's standards on democracy, human rights, and the rule of law in order to guide the development, design, and application of AI. By May 17th, 2024, the *Framework Convention on AI* [*Council of Europe*, 2024] was adopted. This document is of a binding nature for those who sign it and describes the fundamental rights and values that must be upheld throughout AI's life cycle. By recognizing the urgent need for a common

strategy, they advocate for a global legal framework to govern AI activities. This framework should prioritize common principles and rules to preserve shared values and foster responsible innovation, harnessing AI's benefits for their promotion. The focus of their work aims to address challenges across the lifecycle of AI systems by encouraging broader consideration of risks and impacts, encompassing human health, the environment, and socio-economic factors such as employment and labour.

This chapter presents an analysis of the content of the two works of reference (for further arguments in support of this choice, please read [*López Baroni*, 2024, p. 11]). The primary elements of the analysis are illustrated in Fig. 5.1, which delineates four sections of particular interest: scope, measures for safety and governance, and a perspective on autonomy; together with the principal points to be discussed.

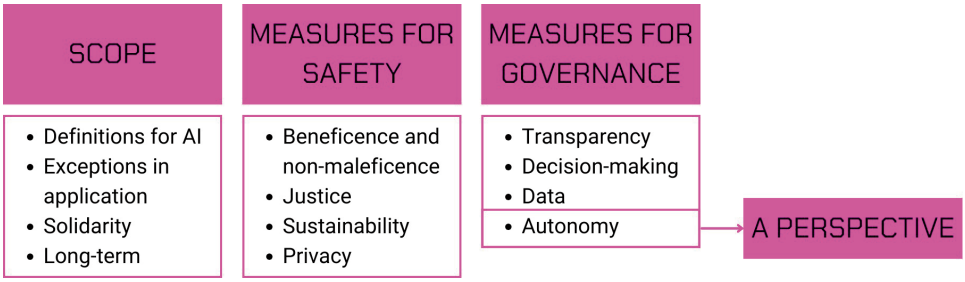


Fig. 5.1. Summary of the key points raised in the analysis of the works of reference produced by UNESCO and the Council of Europe. Each pink box represents a section, and the itemized elements within each section represent the main topics to be addressed.

Both instruments aim to tackle AI systems, but their definitions for AI differ in reach. A clear definition is necessary in order to develop a normative framework that ensures legal certainty[1]. UNESCO aims to target those systems that process data in a way that resembles intelligent behaviour. Despite being a simple definition, it encompasses the different technologies developed under the umbrella term of AI. The most significant drawback is the use of the concept of intelligence, which lacks a broadly accepted definition. The Council of Europe focuses on systems that infer from the received input how to generate an output that may influence physical or virtual environments. This description lacks clarity regarding the complexity and learning capabilities

1. A similar situation arose with another emerging technology: nanotechnology. In that instance, definitions were based on properties, size, or particle count, rather than strict scientific criteria. Evidence of that legal uncertainty is that within a short timeframe, European Union institutions have used varying criteria, which has resulted in notable differences among regulations and directives [*Casado and López Baroni*, 2018b].

that characterize AI systems. To further clarify the type and scope of the impact on environments, the term "influence" could also be described more precisely. Although this definition does accommodate AI-based systems, it also allows identification with many other processes of traditional computational practice, thus presenting a weak framework to specifically target the challenges concerning human rights, democracy, and the rule of law that AI encompasses.

The two works cover all aspects of AI applications, but underline the need for greater prudence in the health care sector. They emphasize the necessity of putting people first and taking domain expertise into account while developing solutions. Because of the sector's sensitivity, there is an understanding of the risks connected with potential data management vulnerabilities. There is also concern about the possible harm that these solutions may cause, particularly in terms of mental health. This is especially important because this sector is considered as requiring the most safeguards, making it the most protectionist domain, necessitating continuous and increased safeguards to preserve safety and ethical standards.

Only the Council of Europe proposes an exception in application, particularly when national security is at stake. Their position highlights certain risks stemming from the inherent imperfections in the judgments made by AI-based systems. Different measures may be taken for the sake of *national security*, being AI-driven surveillance mechanisms one of them. These systems may aim at targeting specific persons of interest but, because of their flawed nature, they can end up focusing on other similar people with significant resemblance. Thus, they have the potential to pose a significant and disproportionate threat to innocent individuals [*Cole*, 2014]. Without delineated limits, surveillance practices can easily encroach upon privacy rights and civil liberties, subjecting individuals to unwarranted scrutiny and suspicion. This unrestricted surveillance not only undermines the fundamental principles of freedom and autonomy but also raises concerns about the misuse of power by authorities [*Snowden*, 2019]. Such unchecked surveillance may not only erode trust in government institutions but also create a chilling effect on individual expression and behaviour. Therefore, establishing robust regulations and safeguards to govern surveillance practices is imperative to uphold the rights and dignity of all individuals within a democratic society.

There is a demand that the way AI systems are approached and understood be based on how they affect human rights, the ability to exercise them, and the well-being of ecosystems and the environment. The pertinent issue at hand is whether the general public understands the broad consequences

of AI systems on human rights. Is there widespread awareness among individuals about their rights? And especially in the fast-developing field of AI technology? This second question gets to the heart of social awareness and consciousness regarding the complex relationship between technology and human rights. There is a significant gap in public understanding of core human rights, with many people being ignorant of how AI systems can specifically affect their rights to privacy, autonomy, and non-discrimination [*Véliz*, 2019]. Bridging this gap needs significant efforts in education and public discourse to ensure that individuals are equipped with the knowledge and awareness needed to protect their rights in the face of technological progress.

While both proposals demand the assessment of the ethical implications posed by AI, only UNESCO details a strategy to do so. UNESCO suggests two different approaches. Both Member States and the private sector share responsibilities for establishing control mechanisms and impact measures. However, when it comes to the adoption and implementation of a regulatory framework, particularly for public authorities, only governments are specifically identified as active actors. While the proposed steps seem appropriate, the allocation of duties may be reconsidered. Many argue that it is the government's primary obligation to protect our rights, acting as a shield against the various interests of private initiatives [*Mazzucato*, 2013].

In the same vein, UNESCO proposes the premise that participants in the field of AI must respect and promote freedom of expression and access to information. However, the vast majority of AI-based systems currently originate from private initiatives, and in the absence of regulation overseeing them, they operate without constraints. One issue that has emerged in this context is related to the automation of censorship, particularly concerning the dissemination of sensitive content [*Williamson*, 2023; *Cobbe*, 2021]. One example is attempting to automatically determine whether a message including a sensitive term is offensive or not, which will ultimately perpetuate the prejudices of the people who define what is an offense. As a result, the entities developing these technologies have implicit power by establishing unregulated measures, potentially reinforcing their own viewpoints. So, is it currently realistic to demand freedom of expression and access to information for all AI initiatives?

ON SOLIDARITY. UNESCO underlines the duty of the most technologically advanced countries towards the least advance to ensure that the benefits of AI technologies are shared to contribute to a fairer world

This book offers both a theoretical and practical contribution grounded in the author's professional experience in artificial intelligence (AI) applied to digital health. It provides a timely and innovative proposal for algorithmic governance rooted in bioethics and anchored in the promotion of human rights and ethical principles internationally recognized. The aim is safeguarding individual autonomy in AI.

The work begins with an analysis of the limitations of machine learning, a branch of AI, identifying gaps and challenges, particularly the misalignment between technological development and ethical imperatives. Moreover it explores the reasons why existing regulation is not being applied. From this perspective, it advocates for an approach that places ethics at the heart of AI design and deployment by providing a critical analysis that includes the review of the state of the art and recent doctrinal contributions. The work analyzes the existing international normative framework to identify the cornerstones where to ground the algorithmic governance proposal and to detect its strengths and weaknesses. By doing so it provides an ethical, legal and societal analysis of the implications of AI applied to research and innovation in health.

Therefore, with a practical vocation to have an impact in digital health and in other research and innovation processess in which human rights could be endangered by AI, the main contribution is to operationalise the principle of autonomy, its understanding, and integration into algorithmic processes.

In sum, the work provides a model of algorithmic governance capable to align societal interests with the development and use of AI in health. This proposal is centered on the monitoring of autonomy throughout the entire algorithmic pipeline, from its inception to completion including the ethical, legal and societal issues. It facilitates continuous assessment of end-users' autonomy with regard to the systems that influence decision-making processes.

ISBN: 978-84-1085-425-3

Bioética
Nanotecnología
Políticas públicas
Biología
Genética
Sociología
Filosofía
Derecho
Antropología
Economía
Biotecnología
Medicina

Observatorio de Bioética y Derecho
Cátedra UNESCO de Bioética
Facultad de Derecho
Universidad de Barcelona
Av. Diagonal, 684, 08034 Barcelona
Telf. (+34) 93 403 45 46
E-mail: obd.ub@ub.edu
Web: www.bioeticayderecho.ub.edu

"Cuestiones bioéticas no resueltas en la evaluación de la investigación e innovación en salud basada en inteligencia artificial, tecnologías genéticas y datos personales (BIOEVAINNOLAW)". IP: Itziar de Lecuona. 1/9/23 - 31/8/27.
Proyecto PID2022-138615OB-I00 financiado por:

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES
Cofinanciado por la Unión Europea
AGENCIA ESTATAL DE INVESTIGACIÓN

UNIVERSITAT DE BARCELONA
unesco Cátedra de Bioètica de la Universitat de Barcelona
Observatori de Bioètica i Dret Universitat de Barcelona
ARANZADI