BioLaw Journal
Rivista di BioDiritto

UNIVERSITÀ DI TRENTO

1 26

Special Issue

a cura di M. Sosa Navarro, E. di Carpegna Brivio, C. Gulotta

Special Issue || AI & Data Security. A Euro-American Dialogue for a New Era of Regulation

The online Journal about law and life sciences

BioLaw Journal – Rivista di BioDiritto
ISSN 2284-4503
Quarterly journal | Rivista trimestrale
Registrazione presso il Tribunale di Trento n. 6 dell'11/04/2014

E-mail: biodiritto@gmail.org **Web**: https://teseo.unitn.it/biolaw | https://www.biodiritto.org/

In collaboration with | In collaborazione con:

Front cover | Copertina*:* Graphic project based on the *Tomb of the Diver*, Paestum, 5th century b.C., on permission no. 1/2026 by the Archaeological Parks of Paestum and Velia - Italian Ministry of Culture | Progetto grafico basato sulla *Tomba del Tuffatore*, Paestum, V sec. a.C., permesso di riproduzione n. 1/2026, Parchi archeologici di Paestum e Velia - Ministero della Cultura
Cover design | Grafica di copertina: Marta Tomasi

Special Issue 1/2026
AI & Data Security: A Euro-American Dialogue for a New Era of Regulation,
edited by | a cura di: Marta Sosa Navarro, Elena di Carpegna Brivio, Carla Gulotta
ISBN 978-88-5541-148-6 | DOI 10.15168/2284-4503-20261S

Published by | Pubblicato da: Università degli Studi di Trento
via Calepina, 14 – I-38122 Trento | casaeditrice@unitn.it | www.unitn.it
March | marzo 2026

Contents

BioLaw Journal – Rivista di BioDiritto
Special Issue 1/2026

Table of Contents – Indice

AI and Data Security. A Euro-American Dialogue for a New Era of Regulation 1
Robert I. Field

ESSAYS – SAGGI

Artificial Intelligence and the Semantics of Change: Narratives, Languages and Values 5
Silvia Salardi

The European Normative Response to the Data Society: From the GDPR to the AI Act 17
Carla Gulotta

Regulatory Tipping Point: Key Lessons from a Divergence in AI Regulation Between the EU and the US 33
Jordan L. Fischer

AI and Personal Data Regulation: From Public Authority Enforcement to Civil Liability Law 53
Erwann Picart-Cartron

Human Dignity and Quantified Self: The Constitutional Challenge of AI 67
Elena di Carpegna Brivio

Artificial Intelligence and Media Freedom: From the 'Brussels' to the 'Strasbourg' Effect? 79
Giovanni Zaccaroni

Workplace Neurosurveillance: Is the Employee's Mental Privacy Protected under International Law? 93
Marta Sosa Navarro

Artificial Intelligence and Credit Scoring: the European Court of Justice Takes Action 115
Francesca Mattassoglio

Digital Technology and the Responsibility of French Legal Professionals 129
Philippe Pierre

Editorial

AI and Data Security: A Euro-American Dialogue for a New Era of Regulation

Robert I. Field

In the short time since Artificial Intelligence (AI) emerged as a broadly available and widely utilized technology, it has begun to transform almost every kind of human activity. As the typewriter replaced handwriting and the pocket calculator replaced arithmetic calculations, AI is altering fundamental cognitive tasks, but it is doing so on a much larger scale. Everyone with a computer can use a number of applications to analyze complex problems based on information gleaned from across the Internet. These applications can not only analyze existing information but also generate entirely new resources based on it, ranging from legal documents to works of fiction to computer code.

Beyond applications for individuals, in the commercial sphere, businesses and professionals are seeing even more consequential uses. In manufacturing, to take one area, AI can detect product defects, program self-driving cars, and predict energy demand. In entertainment, to take another, it can analyze individual preferences, profile consumption behavior, and promote a company's favored content. In medicine, it can help radiologists to read images, oncologists to customize treatments, surgeons to guide robotic procedures, and health systems to manage essential administrative tasks. With its reach into the personal and commercial spheres, AI is touching almost every aspect of our lives.

However, along with its growing array of uses, AI brings risks that are equally profound. AI systems are trained on massive troves of data, some of which may be highly personal or proprietary. They can use the data against the wishes, or even knowledge, of data subjects. Even with information that is legitimately obtained, AI can perform functions that raise serious ethical concerns. Employers can use it to monitor minute aspects of workers' personal behavior and to select job applicants based on characteristics unrelated to objective qualifications. In the United States where employers often bear the burden of providing health insurance for their workers, it can screen job applicants based on predictions of claim costs for them and their family members. Police can use it to profile individuals based on facial features and then monitor the activities of those individuals even if they have never committed a crime. Banks can use it to predict the credit worthiness of loan applicants even if they have never encountered financial difficulties. To compound these concerns, AI analyses of actual and predicted behavior can be not only intrusive but erroneous, leading to devastating consequences for the individuals involved.

Beyond the risks to individuals, governments can use AI in nefarious ways, such as monitoring their citizens' political leanings and activities. They can use it against adversaries to manipulate populations with misinformation. In the near future, data derived from two new highly intrusive technologies, genetic profiling and brain monitoring, may enable governments to use AI to encroach on individual privacy and autonomy in even more fundamental ways.

However, while risks such as these are universal, sensitivity to them is not. It varies with social context, which in turn varies between countries, leading to differences in legal responses. Such disparity is evident in the AI policies of two jurisdictions that have been especially active in its development and oversight: the European Union (EU) and the United States (US).

In May 2025, universities in three countries in the EU and US convened a conference to explore

Editorial

AI's social challenges and legal responses in these two jurisdictions entitled *AI & Data Security: A Euro-American Dialogue for a New Era of Regulation*. The host universities were Università degli Studi di Milano-Bicocca in Italy, Universite de Rennes in France, and Drexel University in the United States. This issue of *BioLaw Journal* presents the perspectives of nine speakers at that conference. Their articles present viewpoints on a range of aspects of AI from theoretical considerations based on legal philosophy to detailed considerations of existing laws and their effects.

Silvia Salardi considers the nature of AI from the viewpoint of analytical legal philosophy, using linguistic analysis to clarify the meaning of concepts. She discerns three different yet interrelated kinds of languages: institutional and political, legal, and ordinary discourse. It is through these linguistic forms that deep transformations of society are expressed and through which societal changes are collectively and individually perceived and experienced. Control over language and the concepts they express thereby forms an important guide to these transformations that is in line with a fundamental rights framework.

Carla Gulotta describes a study that seeks to enhance understanding of the effectiveness of the EU's normative framework in its response to AI. That framework seeks to shape a digital society in which AI systems do not endanger respect for fundamental rights and democratic values. Europe has prioritized these rights and values in its social and legal order since the end of the Second World War. The study will generate observations on the kinds of legal tools that might better shape a rights-oriented society capable of capitalizing on AI without compromising basic values. The study's analysis will emphasize the precautionary principle to inform the innovation process.

Jordan Fischer explores the different attitudes of the EU and US as the two dominant jurisdictions regulating AI, which was a central theme of the conference. She considers how each jurisdiction has approached AI regulation and the legal frameworks they have produced. Their experiences may hold lessons for the next wave of AI development and oversight.

Erwann Picart-Cartron describes administrative procedures before the EU's supervisory authority and the authority's ability to act under liability law and under the EU's two main laws regarding AI and data security: the AI Act and the General Data Protection Regulation (GDPR). On the one hand, the supervisory authority has a somewhat contradictory role as it operates at the intersection of market regulation and the protection of fundamental rights, combining both ex-ante and ex-post powers. On the other hand, liability law must also be considered as a tool for ensuring the enforcement of these regulations, especially given its preventive function. By analyzing these remedies, he highlights the importance of the GDPR as a key component in ensuring the effectiveness of the AI Act and points to ways in which AI requires a new perspective on data protection law.

Elena di Carpegna Brivio examines the relevance of the constitutional concept of dignity in the digital society. Digital technologies are redefining the concept of human personality, employing a quantitative approach that considers human behavior through a statistical lens. The idea of dignity can be a useful tool for creating a new approach to juridical reasoning by drawing a continuous line through a person's physical, psychic, relational, and digital existence, and continuing that line to the AI Act. This reasoning could underly the beginning of a new regulatory philosophy of technological development that would be more anthropocentric.

Editorial

Giovanni Zaccaroni considers media freedom and argues that AI poses both challenges and opportunities for the media in Europe. EU legislation, including the Media Freedom Act, the AI Act, and the Political Advertising Regulation, along with the European Charter and European Convention on Human Rights, seek to protect media pluralism and democracy amid rapid digital transformation. AI's growing role in content creation and distribution raises concerns over media autonomy and editorial independence, and these laws aim to promote transparency, AI literacy, and safeguards against undue influence. He argues that cooperation between the EU and the Council of Europe through frameworks such as the AI Convention is especially important in creating a safe approach to innovation that promotes digital autonomy.

Marta Sosa Navarro analyzes the implications of neurotechnologies used in the workplace for international human rights principles and labor law. She distinguishes between brain-reading devices, which process neural data and implicate the rights to privacy and to freedom of thought, and brain-altering technologies, which have the potential to affect mental integrity. Mapping the international, regional, and International Labor Organization frameworks, she highlights gaps in protection created by fragmented regulation. She argues that the precautionary principle, soft-law instruments, and anticipatory regulation are essential to address these challenges and concludes that safeguarding dignity in the digital workplace requires a principled and proactive governance model to prevent cognitive surveillance and exploitation.

Francesca Mattassoglio describes an important ruling of the European Court of Justice on AI techniques for calculating credit scores. She sees the ruling as especially helpful as a guide, because the judge addressed the activities of specific credit-scoring companies, including Germany's Schufa and the US's Dun & Bradstreet. Similar disputes that require courts to weigh the interests of parties whose scores are calculated against those of the companies that calculate those scores will likely increase over the next few years. In adjudicating them, judges will have to balance the right of individuals to maintain control over their information with the right of companies to use algorithms as an efficient way to analyze it.

Philippe Pierre describes the responsibilities of legal professionals concerning the use of digital technology. He observes that the nature of responsibility for proper use of digital tools is far from clear. In most cases, the use of digital tools will prove neutral for the legal professionals' commitment to responsible conduct, as it leaves unchanged both the paradigm of good professional practice and the protection of clients. Nevertheless, in some circumstances, the digital environment may expand this responsibility in ways that may paradoxically work to the benefit of the practitioner. This will be the case, for example, for the two classic sources of liability for legal professionals: failing to provide advice and committing legal errors.

With this range of perspectives, this issue of BioLaw Journal represents an important resource for understanding the unique legal and societal challenges posed by what may be the most consequential new technology of our time. It is especially valuable in assessing the contrasting approaches of two of the most important jurisdictions in regulating it. Such an analysis is essential if a harmonized global strategy for maximizing the benefits of AI while minimizing the risks is to be achieved.

Essays

Artificial Intelligence and the Semantics of Change: Narratives, Languages and Values

Silvia Salardi*

ABSTRACT: The paper proposes a reflection on the topic of Artificial Intelligence from the perspective of analytical legal philosophy, which uses linguistic analysis to clarify meanings of concepts deployed in three different, albeit interrelated, languages: institutional and political languages, legal language, and ordinary language. It is indeed through these three linguistic vectors that deep transformations of society are narrated. These narratives and their languages shape how societal changes are collectively and individually perceived and experienced. Control over languages and concepts is therefore important to guide these transformations in line with the fundamental rights framework as will be argued in this paper.

KEYWORDS: semantics; analytical philosophy; AI; narratives; anthropomorphism

SUMMARY: 1. Introduction and methodology – 2. Narratives in the Information Society – 2.1. AI as object of narration – 2.2. AI as a narrator agent – 3. Linguistic consistency as a value in the European legal framework: some inconclusive remarks.

1. Introduction and methodology

In this paper I would like to propose a reflection on the topic of Artificial Intelligence (AI) from the perspective of the analytical legal philosophy, which uses linguistic analysis to clarify meanings of concepts with the aim of making language, in particular legal language, appropriate for its purposes in each historical moment for the needs of a given community. The main goal of this paper is indeed to highlight how language and the narratives grounded in specific selected concepts can shape in a subtle and ideological way different, albeit interrelated contexts, in particular civil society, institutions, and law.

The focus of this analysis is on concepts deployed in three different, albeit interrelated, languages:[1] institutional and political languages, legal language, and ordinary language. It is indeed through these

* *Associate Professor of Philosophy of Law and Bioethics, University of Milano-Bicocca. Mail: silvia.salardi@unimib.it.* *This article was subject to a blind peer review process.*

[1] According to Norberto Bobbio, legal philosophy should start from the legal experience to provide tools of analysis of legal concepts and of the distinctive characteristics of law from other normative systems, see N. BOBBIO, *Natura e funzione della filosofia del diritto*, in N. BOBBIO (ed.), *Giusnaturalismo e Positivismo Giuridico*, Milan, 1965, 37-51. Analysis of concepts intersecting different languages has been at the centre of Herbert Hart's reflection on law, H.L.A. HART, *The Concept of Law*, New York, 1961. See also P. TIEDEMANN, *Philosophical foundations of human rights*, Cham, 2023.

three linguistic vectors that deep transformations of society are narrated in the Information Society (IS). The choice of this linguistic and conceptual analysis is justified by the transformative and shaping role that narratives play through their languages in framing the common human understanding of a given phenomenon.

Put it differently, narratives and their languages shape how societal changes are collectively and individually perceived and experienced. And both the perception and the experience impact the modulation and timing of the process of acceptance, adaptation, and normalization of the phenomenon driving the changes of our time. Thus, the language, in which a narrative is grounded, is a powerful tool that needs to be carefully managed if transparency in communication, especially at the institutional level, is considered a value to pursue.

The relevance of such an attentive attitude becomes evident when we consider the common reaction of the public generally provoked by deep transformations of society: a mixed reaction, which tends to be polarized between demonization and uncritical enthusiasm. None of these reactions permits dialogue and understanding based on reason, knowledge, and facts, because they are too emotionally charged.

Instead, to exercise control over the deep transformative forces of our time we need to act and react based on knowledge and understanding of facts, and to argue for or against these changes based on rational arguments. This means being aware that the phenomenon we face and discuss is socially situated, and therefore the key to its understanding is how it is defined in that social context.

For this reason, control over narratives and their languages is of the utmost importance to avoid or at least limit misunderstanding, miscalculation, and mistakes. What does it mean to exercise control over language in this article? It means that those in charge of promoting the *human-centric vision of AI*, which characterizes the European Union approach, integrate control over the selection process of the key concepts that build the structure of that vision in the early stages of the elaboration of institutional normative documents (both soft law and hard law). In doing so, they grant that linguistic consistency is maintained throughout the process of the strategy's elaboration both within any single normative document (intra-documental linguistic and conceptual consistency) but also among the various institutional documents (extra-documental linguistic and conceptual consistency) that tackle the topic from different perspectives, though being part of the same value-laden vision. Thus, the goal is to grant consistency in the linguistic interrelatedness of strategic documents based on the inter-definability of some primitive concepts. This linguistic consistency is the basis for achieving normative consistency[2] as much as possible. Indeed, one basic challenge while preparing an institutional strategic plan on a specific issue is to carefully select the key notions that build the conceptual frame in which different documents pertaining to that strategy are combined. The more complex the strategy is higher is the risk that conceptual and normative consistency[3] is not maintained, with relevant implications for its effectiveness and trustworthiness.

[2] L. FERRAJOLI, *Dei diritti e delle garanzie. Conversazione con Mauro Barberis,* Bologna, 2013, 22. The author underlines that "normativity is the stronger and binding the more unambiguous and rigorous is the normative language", the original sentence is "La normatività è tanto più forte e vincolante quanto più univoco e rigoroso è il linguaggio normativo".

[3] Conceptual consistency is not to be confused with precision. Indeed, whereas consistency is an essential value for any system that wants to convey a specific message, precision is a relative value whose pursuit needs a case-by-case evaluation.

This problem may be due to different factors, but linguistic indeterminacy and vagueness, especially when they are unintentional, are relevant causal determinants of it. Linguistic control as described begins with soft law documents elaborated to facilitate the consensus on legally binding rules that may follow. For this reason, the language adopted in soft law documents, such as ethical guidelines, should undergo deep scrutiny to eliminate or drastically limit ambiguous and vague concepts borrowed from the ordinary language, if not properly redefined. These soft law documents usually represent a first step in the drafting of legally binding rules. And these rules need to be based on conceptual clarity and logical rigor in order to avoid creative or arbitrary interpretations that may result in inconsistencies with norms and with fundamental constitutional principles.

In this sense, the linguistic consistency, which is useful to maintain conceptual and normative consistency, is a value to be preserved in the construction of the EU vision on technological development, so that EU citizens, as final users of the AI technology, have clear reference points for their choices and can understand the value-laden EU approach. Indeed, among the many threats related to the AI development, some are realistic, and others derive from a distorted perception. This gap between realistic and unrealistic threats prominently depends on the mainstream narrative around AI, which has its roots in the past century: both in science fiction and in the same notion of AI.[4]

In what follows, I will identify some criticalities concerning the mainstream narrative around AI, that is, when AI is the *object of narration*. However, as we are faced with a peculiar technology that unlike traditional technologies is also able to generate its own narratives (generative AI), to manipulate languages, and consequently to shape perceptions and experiences of users, I will also discuss AI as a *narrator agent*.

This double face of AI (object and subject of narration) renders this technology uniquely pervasive to the extent that it can be classified as a new *formant* of human existence along with traditional ones, such as religion, culture, science, etc.[5]

This unprecedented scenario needs to be investigated with regards to its new and unforeseeable outcomes on human existence within what I suggest terming the *semantics of change*.[6]

2. Narratives in the Information Society

Before dealing with the twofold nature of the narratives concerning AI, let us characterize them and put them in context. In the current model of society (IS), information is the raw material around which human activities and their organization are structured. This information is assembled within narratives around different topics that accompany societal changes brought about by AI.

[4] As is well known, the notion was coined during the Dartmouth Summer Project on Artificial Intelligence in 1956.

[5] On this point see S. JASANOFF *et al*., *CRISPR Democracy. Gene Editing and the Need for Inclusive Deliberation,* in *Issues in Science and Technology*, 32, 1, 2015, 25-32, where the authors observe that "Science and technology not only improve lives but shape our expectations, and eventually our experiences, of how lives ought to be lived".

[6] This notion is part of the title of my book on this topic: S. SALARDI, *Intelligenza Artificiale e Semantica del Cambiamento: Una Lettura Critica*, Torino, 2023, in which the reader can find more detailed reflections on this topic. The expression *semantics of change* refers to the influence that a skilful use of language through specific operations on the meaning of its socially and culturally situated concepts can exert on societal transformations.

These narratives are of two kinds: descriptive and prescriptive-normative. Descriptive narratives describe typologies of AI, the state of the art of the technological developments, what are possible applications, and how they function. These narratives primarily aim to inform the public. Descriptive narratives may be semantically neutral or use terms and notions with a persuasive connotation, as in advertising, for instance. Unlike descriptive narratives, prescriptive-normative narratives are never semantically neutral. They are axiologically grounded and are used to convey a position, create or affirm legitimacy, or justify a means to an end. Through these narratives, particular policies and their directions are prioritized and justified.

The linguistic vectors of these narratives are different, albeit intertwined: ordinary language, institutional-political languages,[7] and legal language.

The link between these vectors is represented by a taxonomy of concepts which can be ubiquitously deployed in the three languages. They are concepts of the ordinary language, borrowed from legal language as well as from the institutional-political one.[8] As these notions are part of the ordinary language's vocabulary, they carry different layers of meaning, 'socially determined' vagueness,[9] ambiguity, and indeterminacy,[10] which are all features of ordinary language. These constitutive aspects of the ordinary language are not critical *per se.* Rather, they become problematic if transposed into technical and artificial languages without mechanisms in place to limit or eliminate them in order to allow these languages to achieve their goals. Legal language has, for instance, the main aim to provide legal certainty and to be a guidance for behaviours consistent with shared values. If notions deployed in this language are too emotionally exposed or too indeterminate to be reframed by means of re-definitions, this will impact the legal effectiveness of rules and the ability of the law to guide behaviours.[11]

In other words, linguistic control over key concepts borrowed from ordinary language plays an important role in limiting the ideological instrumentalization of that narrative, for instance at the legal

[7] Although there are differences in the institutional language (*eurojargon*) and in the political one, they often overlap, and therefore, for the purposes of this article, they will not be distinguished.

[8] Some examples are responsibility, autonomy, (un)predictability, etc.

[9] Vagueness is a characteristic of meaning and is a matter of degree referring to all notions, including notions of artificial languages. However, in some cases we should ask whether the problem we are facing is vagueness as just described or whether we are dealing with a meaning of a notion that we do not consider aligned with the system of values in which the notion is deployed. Some notions are vague because they express a value judgment whose application requirements are not even partially determinable unless reference is made to variable judgment parameters and to the changing typologies of social morality and customs. C. LUZZATI, *La vaghezza delle norme. Un'analisi del linguaggio giuridico*, Milano, 1990.

[10] When discussing linguistic problems, it is important to maintain the distinction between indeterminacy and vagueness. The first notion may include vagueness, but it is broader and refers not just to situations where it is unclear whether the word can be used or not (penumbral zone or fringe), which are usually defined with the term vagueness, instead it may refer to cases that are not determined at all. See P. VAN INWAGEN, *Indeterminacy and Vagueness: Logic and Metaphysics*, in *European Journal for Philosophy of Religion* 1, 2, 2009, 1-19.

[11] As was clearly explained, "the question of semantic rigor of legal language is prodromic to the upholding of the rule of law and of democracy", L. FERRAJOLI, *Dei diritti e delle garanzie,* cit., 22. The original sentence is "La questione del rigore semantico del linguaggio legale è dunque pregiudiziale alla stessa tenuta dello Stato di diritto e della democrazia".

level, and represents also a propaedeutic operation to examine how dominant narratives emerge and to unveil underpinning power relationships.
In the next paragraph, I will discuss AI as object of narration and highlight how a specific dominant narrative around AI has been reiterated over time, and why this is a matter of concern in legal terms.

2.1. AI as object of narration

Since its birth in 1956,[12] AI has been at the centre of science fiction and literature that have contributed to shaping the mainstream narrative on AI. This narrative is basically built upon a *dialectic of opposites*[13] without synthesis of reconciliation, which aims to define artificial intelligence *per relationem* with human intelligence.[14] The immediate result of such a comparative contrast is to highlight the extraordinary abilities of AI to the detriment of the natural limits of humans. Consequently, this narrative is trapped in emotionally charged modalities of presentation of AI versus humans, which do not allow rational approaches to the discussion around this topic. In this scenario, the engaged race between AI and humans always results in the AI's victory over humans.
The limits of this narrative and of the AI notion itself have been institutionally recognized in 2019, when the *High-Level Expert Group on AI* clearly acknowledged the importance of redefining AI following a semantically neutral approach. The Group proposed an AI definition useful to "avoid misunderstandings, to achieve a shared common knowledge of AI that can be fruitfully used also by non-AI experts, and to provide useful details that can be used in the discussion on both the AI ethics guidelines and the AI policies recommendation".[15]
This has been an important linguistic operation of clarification and re-definition, able to emancipate the concept of AI from ideologically connoted meanings as well as from historically rooted hypostatizations.
The same operation has, however, not been conducted consistently in the process of developing the European strategic approach to AI, which is based on building an *ecosystem of trust*[16] for a human-centric AI. As a matter of fact, other notions deployed in the mainstream narrative and in the institutional-political language, as well as in the legal language, have not been deeply scrutinized, thereby failing to prevent ideological stances that defend a reductionist view of humans in their relationship with AI.
In what follows, I will focus on three of the key notions ubiquitously deployed in the three above-mentioned languages. The skilful use of these notions has ideologically shaped the mainstream narrative

[12] The notion AI has been coined during the Dartmouth College Project on Artificial Intelligence. Propaedeutic to the elaboration of this notion were previous scientific studies such as Alan Turing's essay on *Computing Machinery and Intelligence* published in *Mind,* LIX, 236, 1950, 433-460.
[13] S. SALARDI, *Intelligenza Artificiale e Semantica del Cambiamento: Una Lettura Critica*, cit.
[14] The title of the book by L. ALEXANDRE, *La guerra delle intelligenze. Intelligenza artificiale contro intelligenza umana,* Torino, 2018, is very telling in this sense.
[15] High-Level Expert Group on Artificial Intelligence, *A Definition of AI: Main Capabilities and Disciplines,* Brussels, European Commission, 2019, 1. The definition proposed explains that "AI refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals" and by AI system is intended "any AI-based component, software and/or hardware. Indeed, usually AI systems are embedded as components of larger systems, rather than stand-alone systems", 86-87.
[16] European Commission, *White Paper on Artificial Intelligence – A European Approach to Excellence and Trust*, COM(2020) 65 final, 2020, 9.

of the past decades in the European context, influencing also the institutional decision-making process, as we will see in what follows.

The notions are revolution, personhood, and trust.[17]

Let us begin with *revolution*. This is the notion around which changes brought about by AI have been long narrated through media and literature. This is a very critical concept, as it has stratified different layers of meanings, and when it is used without redefinition, it can serve different and contradicting purposes.

What does it mean that AI is revolutionary? The first meaning that is evoked by this notion is its *strongest meaning*, the one that refers to a radical change of paradigm, where the traditional epistemological schemes are completely overruled, as was the case with the scientific revolution.

When this notion is used to describe deep societal transformations like those brought about by AI, the strong meaning is the first that usually comes into the mind of laymen. This fact is not *per se* positive or negative. It is the *ex post* axiological evaluation of this use that may reveal a problematic hidden aim in the communication process, that is, that the senders use the word without defining it to make clear what they mean with the term, and in doing so, they do not put the receiver in the position to critically approach the societal changes being discussed. Indeed, in this scenario, the receivers are the passive recipients in a communication process following a one-way direction with a predefined purpose. When this communication strategy is adopted by governing institutions in democratic contexts framed by fundamental rights, it may be the indicator that the pressure of private actors or lobbies on the decision-making of those institutions is strong and is exerted to divert the legislation process in order to safeguard private interests.

In fact, when the sender uses the concept of revolution without re-definition, she evokes the strong meaning that influences the public perception of the phenomenon in a very specific way. What is surreptitiously suggested is that this phenomenon has not been planned and governed, therefore it is largely inevitable.[18] Following this semantic strategy allows the feeding of deterministic views of technology, which can ideologically shape the relationship between society, individuals, and technology[19] in two related ways. On the one side, it is maintained that technology autonomously shapes and guides society and cultural values. On the other side, technology is assumed to be neutral about the context in which it is developed, and therefore it predetermines the inevitable societal path.

However, the history of AI leads us to a different conclusion than the one referred to with the strong meaning of revolution. From the very beginning, the AI development has followed a precise direction, guided by private and rich stakeholders through the past century. It is a revolution, but in *a weakest sense*, as synonymous of transformative innovation. To highlight this weak nuance in the meaning of the notion revolution unveils that the AI innovation is a process that has been planned and governed, and therefore it is still governable in its further evolution. It is a matter of political and societal will to define

[17] I have analysed further concepts of the conceptual taxonomy building the mainstream narrative and impacting institutions and law in the book *Intelligenza Artificiale e Semantica del Cambiamento: Una Lettura Critica*, cit.

[18] S. ZUBOFF, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, New York City, 2019.

[19] R.R. KLINE, *Technological Determinism, in International Encyclopaedia of the Social & Behavioural Sciences (Second edition)*, Amsterdam, 2015, 109-112.

the directions of this evolution, whether it should lead to *algorithmic governmentality*[20] or remain under the control of politics, law, and social norms. This conclusion is of the utmost importance when one discusses the role of the law in defining the limits of this transformative innovation and in providing rules for tech-companies which possess both the economic and technological power, which in turn defines timing and characteristics of this innovation through the mainstream narrative.

If revolution is the key concept around which changes brought about by the advent of AI have been narrated so far, the concept of *personhood* is the pivot around which the anthropomorphising process took place.[21] This process started with the same notion of AI and has been strengthened through science fiction and literature, and this process can now be fulfilled and even legitimized through the legal language and its categories.

Personhood is indeed a central notion in modern constitutionalism, as it refers to the holder of human rights, and it is also the basic value standard for dignity.[22] Starting from 2017, an institutional and legal discussion in the European Union has begun with focus on expanding human rights protection to non-human subjects, such as autonomous, self-learning, and unpredictable robots.[23] This discussion has been strongly affected by the absolutizing and objectivistic hypostatization of AI in the ordinary language that the notion of legal personhood could have contributed to enhancing and enforcing. Indeed, recognition of legal personhood to AI has both a high symbolic relevance and a value-laden dimension. As a matter of fact, human rights are the bridge between ethics and law. They represent the legal device useful to frame the impact of technology on human existence by providing a process of agenda setting of priorities, sphere of influence, and responsibility that puts the person at the centre with her needs and expectations. Being a person in legal terms means to be a rights-holder deserving respect for their dignity. In other words, to be legally qualified as a person gives a trump card, as it allows to be included in policies granting the distribution of benefits. Recognition of legal personhood to AI systems or robots is therefore a very delicate question as it can both strengthen and even legitimize the tendency to anthropomorphism. This notion refers to the attribution of human qualities, traits, emotions or intentions to non-human entities. Anthropomorphism is a natural tendency of humans in their relationship to non-human entities, and its degree during interaction with robots or AI systems vary depending on many factors.[24] Despite being part of the normal interactions that humans develop with non-human entities, the negative implications of such phenomenon may be very serious and are

[20] A. ROUVROY, T. BERNS, L. CAREY-LIBBRECHT, *Algorithmic governmentability and prospects of emancipation*, in *Réseaux*, 177, 2013, 163-196.

[21] This process aims to attribute human-like characteristics and abilities to non-human entities.

[22] On the principle of dignity as the basis for human rights see P. TIEDEMANN, *Philosophical foundations of human rights*, cit.

[23] In the European Union, the proposal to recognize *e-personality* to robots went in this direction. Although the intent was to use this category in legal-technical terms, the notion was trapped into a wider discussion about the constitutional value of this category. See the European Parliament Resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics, www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html (last visited 10/02/2026).

[24] R. KÜHNE, J. PETER, C. DE JONG, A. BARCO, *How does Children's Anthropomorphism of a Social Robot Develop Over Time? A Six-Wave Panel Study*, in *International Journal of Social Robotics*, 16, 2024, 1665-1679.

known as *dishonest anthropomorphism*.[25] To contrast such negative side effects of human tendency to anthropomorphize AI, control over the use of language in media as well as in technical fields such as the law can avoid confusion between the phenomenon of anthropomorphism (descriptive level) and its legitimization through the *performative* language of the law (normative level).

This control has neither been constant nor consistent in the EU implementation of legal norms concerning AI. The evidence of this observation comes from the analysis of another concept building the conceptual taxonomy of the AI debate and being a relevant factor in the anthropomorphising tendency: trust.

The EU institutional-political narrative frames the relationship between users and AI products within an *ecosystem of trust*, based on a *trustworthy* AI. The first comprehensive document on this topic is the 2019 *Ethics Guidelines for a Trustworthy AI*. In the guidelines' glossary a justification is given for deploying the notion of trust: "While 'Trust' is usually not a property ascribed to machines, this document aims to stress the importance of being able to trust not only in the fact that AI systems are legally compliant, ethically adherent and robust, but also that such trust can be ascribed to all people and processes involved in the AI system's life cycle".[26]

The proposed definition of the notion of trust indirectly assumes that trust is a property that can be ascribed or not ascribed to certain types of entities. However, trust is neither a property of an entity (human or not human) nor is it the relationship itself. Rather, it is a property of that relationship.

This means that, for a relationship to be considered trustworthy, certain conditions must be met, conditions that pertain to the relationship itself, not to the entities involved. Reciprocity is one of this basic features. Reciprocity does not *per se* require an affective bond (a feeling of liking for a person) between the entities that are part of the trustworthy relationship. Indeed, a trustworthy relationship can be developed between an expert such as a physician or a lawyer, and a patient or a client. In these cases, the key element of reciprocity is the competence of the expert and the positively expected outcomes of that competence on the patient's or client's conditions. As these expert profiles are the facework of the system in which they work, they also represent a key to trust those systems, for instance, the healthcare system or the judicial system. How can we translate this scenario into the discussion on AI? What does it mean to ascribe trust to *all people and processes involved in the AI system's life cycle*? Who are the facework of the AI systems? Producers, programmers, tech-companies, or who?

What kind of relationship can the user build with them? Are these stakeholders available for a face-to-face exchange with the final user? Is this exchange not only practically impossible but also conceptually nonsensical? Is it, therefore, appropriate to term it a *trustworthy relationship*?

Concerns about this notion had already arisen by Thomas Metzinger, who was a member of the group of experts who elaborated the *Ethics Guidelines for Trustworthy AI.* He observed that "Artificial Intelligence

[25] B. LEONG, E. SELINGER, *Robot Eyes Wide Shut: Understanding Dishonest Anthropomorphism*, In *Proceedings of the Association for Computing Machinery's Conference on Fairness, Accountability, and Transparency,* 2019, 299-308.

[26] High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for a Trustworthy AI,* 2019, 40, op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1 (last visited 10/02/2026).

cannot be trustworthy. The Trustworthy AI story is a marketing narrative invented by industry, a bedtime story for tomorrow's customers".[27]

Despite these concerns, the expression has been deployed in normative documents building the AI vision in the European context, even when the more neutral synonym is contextually deployed.[28]

In light of the previous considerations, we can affirm that control over selection of notions and (re)definition of concepts has not followed a consistent path in building the political and legal strategies regarding AI in Europe. Despite positive results in this context, such as the effort to be the pioneer in AI regulation worldwide, the inconsistency in institutional and legal language, caused by the absence of consistent control over the key concepts of the institutional narrative, can result in at least two different problems: on the one side, further feeding dishonest anthropomorphism, and on the other side, undermining trust in the EU human-centric protection system.

In sum, both the institutional-political language and the legal language are strongly, albeit often inadvertently, affected by the absolutizing and objectivistic hypostatizations of AI. This is not limited to the general public discussion on AI, but it also impacts the narrative around data, which are the constitutive elements of AI functioning.

In the mainstream narrative, data are often presented through an *objectivistic rhetoric*.[29]

According to this rhetoric, data have a pre-social origin, where subjectivity is present only in the phase of analysis and interpretation and not in the process of collection, or even completely ruled out. Following this representation, what is suggested is that AI is able to give objective, definitive, and complete answers on what constitutes reality and its components. However, this cannot be true, not least because data used by algorithms are not representative of the present situation, as they have been collected in a given moment of the past and are in constant need of being updated. How frequently and carefully data are updated is a key to understand what portion of the assumed reality is represented at a given point in time by data elaborated by AI and to acknowledge how faithful to reality that representation is. And this operation remains quite obscure for most of the end users, even for those who work in technical contexts, such as, for instance, medicine or scientific research, where the availability of updated datasets is of the utmost importance.

2.2. AI as a narrator agent

The traditional way of narrating advances in the technological field is no longer the unique context to investigate in order to discuss the *semantics of change.* Indeed, AI being an object of narration is only half the story.

Unlike traditional technologies, AI is very peculiar, and its development in the field of generative AI makes it a uniquely disruptive kind of technology. Not only is it pervasive, but it can answer questions in a human-like way and generate its own content for a variety of texts matching thousands of millions of

[27] T. METZINGER, *Ethics washing made in Europe,* in *Tagesspiegel*, April 8th, 2019.

[28] Let us take the example of the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law. Art. 12 titles *Reliability*, but in the text of the article *trust* is the notion deployed.

[29] D. BALAZKA, D. RODIGHIERO, *Big Data and the Little Big Bang: An Epistemological (R)evolution,* in *Frontiers in Big Data,* 3, 31*,* 2020, 1-10.

information on which it is trained. We have therefore reached a point in the technological evolution where traditional control over languages and narratives is no longer enough. We face an unprecedented challenge whose contours are far from being clearly delineated.

What is sure is that language(s) is the key to societal transformations, but the exclusive generator of these languages is no longer the human being.

As the Secretary-General of the United Nations has remarked in his speech during the *Security Council meeting on AI and the maintenance of international peace and security*, held in New York on December 2024, threats to peace and security derive also by the ability of AI of creating "highly realistic content that can spread instantly across online platforms – manipulating opinion, threating information integrity and making truth indistinguishable from outright lies. Deep fakes could trigger diplomatic crises, incite unrest and undermine the very foundations of societies".[30]

The envisioned problems concern all the fields in which generative AI can be used. The same Secretary-General of the United Nations noted on May 3, 2025 during the World Press Freedom Day,[31] that if we assume that freedom for people depends on freedom of the press "[…] Artificial intelligence can support freedom of expression – or stifle it. Biased algorithms, outright lies, and hate speech are landmines on the information superhighway. Accurate, verifiable, fact-based information is the best tool to defuse them. […] AI must be shaped in a way that is consistent with human rights and puts facts first".

We can extend these considerations to include all contexts in which generative AI is applicable. Another example of a context in which generative AI should be carefully monitored when generating texts is the context of health and medicine. As the publication titled *Generative Artificial Intelligence in Health and Medicine* by the National Academy of Medicine reports "[…] applications of GenAI in health and biomedicine raise unique risks. These include information inaccuracy relevant to medical decision making due to so-called hallucinations or confabulations; inequitable access, utility, and applicability of LLMs in lower resourced environments; and the perpetuation of biases present in training data or introduced by AI engineers".[32] These problems arise from algorithmic brittleness that may arise "from an algorithm's inability to effectively generalize across datasets or adapt to environmental changes".[33]

These technical problems can have a severe ethical and practical impact on the physician-patient relationship as incorrect information may lead to serious negative outcomes in at least two ways. On the one side, incorrect information may lead to incorrect diagnosis, which will impact patient's health. This is a practical as well as an ethical issue. On the other side, incorrect information generated by AI may have the paradoxical consequence of undermining the trust on which the physician-patient relationship is based. Therefore, AI may be the causal determinant in undermining trust between humans.

In light of the observations expressed so far, it emerges that the problems caused by AI as narrator agent may be very complex and difficult to manage as they concern both empirical aspects (for instance, wrong diagnosis and impacts on patient's health), ethical aspects (autonomy and consent), and

[30] www.un.org/sg/en/content/sg/statement/2024-12-19/secretary-generals-remarks-the-security-council-artificial-intelligence-bilingual-delivered (last visited 10/02/2026).

[31] www.un.org/sg/en/content/sg/statement/2025-05-03/secretary-generals-message-world-press-freedom-day-scroll-down-for-french-version (last visited 10/02/2026).

[32] National Academy of Medicine, *Generative Artificial Intelligence in Health and Medicine. Opportunities and Responsibilities for Transformative Innovation*, 2025, 1.

[33] Ivi, 13.

relational ones (for instance, the trust on which a relationship is based). To grant an effective control over the linguistic texture, the sources on which AI is trained should indeed be so in depth scrutinized that the operation would result in a huge effort – according to some, a disproportionate effort – compared with the positive results that can be achieved and the huge financial gain that companies can obtain in the short-term. In this context, linguistic control may be the indispensable foundation upon which the human-centric vision depends; but that does not stop those having financial and economic interests trying to dispense with it.

3. Linguistic consistency as a value in the European legal framework: some inconclusive remarks

Despite the complexity and the difficulties of a serious engagement in control over consistency of language both when AI is an object of narration or when it is a narrator agent, the pursue of this control must not be labelled as a futile, terminological dispute. On the contrary, this kind of control over linguistic consistency is a precondition assumed within the legal framework based on fundamental rights to grant respect for and protection of those rights. These value-laden framework can properly work and protect individuals from technological threats if axiological antinomies are under control or at least if there are effective self-healing mechanisms to solve them. As axiological antinomies may arise from linguistic inconsistencies between the assumed definition of a principle and its incoherent transposition into technical norms due to unintentionally or intentionally hidden vagueness, ambiguity, and indeterminacy of key concepts, it goes without saying that not healing these antinomies has interrelated side effects: on the one side, the persistence of axiological antinomies undermine the ability of fundamental rights to maintain the human-centric vision in guiding technological advances, and, on the other side, their presence causes the unwanted result of leaving the control of technological development in the hands of technology (technology as a regulating agent). [34]

All the critical aspects mentioned so far allow to conclude that constant reflection on language, its concepts, its different contexts of use (technical, political, legal, ordinary) is of paramount importance if the institutionally declared goal is to fulfil the human centric vision.

In this sense, control over language through the different tools we have at our disposal is a precondition to restore language "as a good means of expression, communication and guidance, good not in terms of internal criteria of its functioning, but good for humans with their current attitudes, needs, projects in the circumstances and situations in which they find themselves".[35] In order to achieve this objective, the inclusion of legal analytical philosophers, who are expert in linguistic analysis, in the expert groups which elaborate guidelines and regulations at the institutional level, could be an added value to the exchange of ideas, in order to make the language of the different, albeit interrelated, institutional

[34] S. SALARDI, *Tecnologie per l'etica del futuro o etica per le tecnologie del futuro?*, in *Humanidades & Technologia Em Revista*, 40, 1, 2023, 5-17.

[35] U. SCARPELLI, *Filosofia analitica, norme e valori,* Milano, 1962, 17. The translation is made by the author of this article. The original excerpt is the following "[…] il linguaggio è tornato a essere, per l'espressione, la comunicazione e l'orientamento un buon mezzo: buono secondo i criteri interni al linguaggio, buono per l'uomo con i suoi attuali atteggiamenti, bisogni e progetti, nelle circostanze in cui ora si trova".

documents dealing with AI compliant with the ethical vision that the European institutions promise to promote. But the inclusion of such experts should be provided also when designing generative AI tools.
The importance of the contribution of the analytical linguistic philosophy was deftly clarified by Uberto Scarpelli some decades ago. He was reflecting on science, but I think that his considerations are still valuable and can be extended beyond science to include the current technological development. Therefore, I conclude my reflection with his words:

> the conceptual framework offered by the analytic philosophy to arrange our ideas and guide our behaviours is the most adequate for the current society, in which the horizon of science is constantly widening, and the explicit or implicit values promoted by science appear stronger and more extended. Indeed, legal analytical philosophy with its general focus on determining and clarifying language reestablishes consistency between attitudes in the scientific field and in the other disciplinary contexts, limiting the paradox of those that in science accept and follow the values of understanding, coherency, and control and outside science yield to the temptation of opposites values.[36]

[36] U. SCARPELLI, *Filosofia analitica, norme e valori*, cit., 33. The proposed translation is made by the author of this article. The original excerpt is the following "[…] il telaio concettuale fornito dalla filosofia analitica per il riordinamento delle nostre idee e dei nostri atteggiamenti sia il più appropriato alla civiltà del nostro tempo, in cui fra i vari orizzonti sempre più si allarga e domina quello della scienza e i valori espliciti o impliciti nella scienza appariscono più saldi ed estesi. La filosofia analitica, infatti, con il suo lavoro di generale determinazione e chiarificazione del linguaggio ristabilisce una coerenza fra gli atteggiamenti nel campo della scienza e gli atteggiamenti negli altri campi, contro il paradosso di chi accetta e persegue i valori della comprensione, della coerenza e del controllo e fuori dalla scienza cede a valori opposti".

Essays

The European Normative Response to the Data Society: From the GDPR to the AI Act

Carla Gulotta*

ABSTRACT: This study aims to improve our understanding of how effective the European Union's normative framework is at achieving its declared objective of shaping a digital society in which the adoption of AI systems does not endanger widespread respect for fundamental rights and democratic values. After providing an overview of the main features of the EU's legal framework on AI, the study will conclude with recommendations on how to better shape a rights-oriented society that capitalizes on AI without compromising the EU's values. The study will also propose the broader application of the precautionary principle to inform the innovation process, not just as an interpretative tool.

KEYWORDS: artificial intelligence; fundamental rights; fundamental rights impact assessment - FRIA; precautionary principle

SUMMARY: 1. Introduction – 2. The main components of the European AI ecosystem: an integrated normative network – 2.1. The empowerment of EU society – 2.2. The promotion of a shared approach to AI at international level – 3. The claim to a 'human-centered' regulation of AI: a critical assessment – 4. Conclusions: *plaidoyer* for a strengthened precautionary approach.

1. Introduction

Europe's focus on the circulation of data has been pioneering and marked from the outset – that we may set in 1981, when the of Council of Europe Convention 108 for the protection of individuals on the processing of personal data was adopted[1] – by the objective of ensuring that the process of progressive digitalization of society, correctly perceived as inevitable, develops in a manner consistent with the protection of fundamental rights, democratic values and the rule of law. It is since 2018, when the General Data Protection Regulation (GDPR)[2] came into force, that

* *Associate professor, Department of Business and Law, University Milano-Bicocca, Milan, Italy. Mail: carla.gulotta@unimib.it.* *This article was subject to a blind peer review process.*

[1] Council of Europe, *Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data* (ETS No. 108), Strasbourg 28/01/1981. For historical, philosophical, and economic reasons behind the sensitivity shown in Europe to the issue of personal data protection, see G. DELLA MORTE, *La regolazione dell'AI: profili internazionalistici*, in O. POLLICINO, F. DONATI, G. FINOCCHIARO, F. PAOLUCCI, *La disciplina dell'intelligenza artificiale*, Milano, 2025, 67-82.

[2] *Regulation (EU) 2016/679 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, and repealing Directive 95/46/EC (General Data Protection Regulation)*, O.J. L 119, 4.5.2016, 1.

the European Union has been considering (with the Communication "*AI for Europe*")[3] how to apply this approach to artificial intelligence.

Focusing on how data must be managed, the Union builds on the experience of GDPR. But while the GDPR is intended to provide protection for personal data, the relationship between AI and data is more complex: generative AI requires very large amounts of data to operate. So, the problem shifts from the protection of personal data – which remains relevant, so much so that one of the regulatory sources of the AI Act[4] is Article 16 of the Treaty on the Functioning of the European Union (TFEU)[5] – to the protection of the rights of the individual and the interests of society that may be put at risk by the processing also of different types of data, collected or inferred (non-personal, economic, commercial...), including metadata.

The increase in the size of data sets and in computational power makes it possible for AI systems to detect and make inferences capable of conditioning people's private lives and driving society out of the paths of free and democratic political choices. The potential hazard widens from the risk of infringement of the right to privacy, to a whole host of human rights that can be violated, ranging from the right to freedom of thought, to the right to health. Hence, the tension between protecting against AI-related risks, and promoting technological innovation.

The uncertainty about the potential and timing of AI's development – recently confirmed by some of the world's most experienced scientists in this field in the International AI Safety Report of January 2025 – conveys a call to caution.[6] The legal instrument marked by the utmost prudence and efficacy in protecting fundamental rights – including the right to privacy – should be identified in the precautionary principle.

Where scientific data do not permit a complete evaluation of the risk, recourse to this principle may, for example, be used to stop the deployment of AI systems such as deepfakes technology on platforms

---

[3] *Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Artificial Intelligence for Europe*, COM (2018) 237 final, Brussels, 25/4/2018.

[4] *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 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)*, *O.J.* L series, 12.7.2024. Among the many commentaries on the act, see: C. CASONATO, B. MARCHETTI, *Prime osservazioni sulla proposta di regolamento dell'Unione europea in materia di intelligenza artificiale*, in *Biolaw Journal*, 3/21, 2021; J. LAUX, S. WACHTER, B. MITTELSTADT, *Trustworthy artificial intelligence and the European Union AI Act: On the conflation of trustworthiness and acceptability of risk*, in *Regulation & Governance*, 2023, 1 ff; L. COTINO HUESO, D.U. GALETTA *The European Union Artificial Intelligence Act: A Systematic Commentary*, Napoli, 2025.

[5] Article 16 – "1. Everyone has the right to the protection of personal data concerning them. 2. The European Parliament and the Council, acting in accordance with the ordinary legislative procedure, shall lay down the rules relating to the protection of individuals with regard to the processing of personal data by Union institutions, bodies, offices and agencies, and by the Member States when carrying out activities which fall within the scope of Union law, and the rules relating to the free movement of such data. Compliance with these rules shall be subject to the control of independent authorities. (…)".

[6] In his introduction to the January 2025 report, Professor J. Bengio, who chairs the group of 96 international AI experts that conducted the study, warned that new risks had emerged since the publication of the May 2024 interim report. This illustrates the rapid pace at which AI technology is developing. The updated text of the report, from October 2025, can be found at: https://internationalaisafetyreport.org/.

accessed by minors, as the psychological impact that children would suffer, especially when deepfakes are used for cyberbullying or gender-based violence, is presently unknown. The legal constraints provided under the AI Act – transparency and labeling for AI-generated content – might be not sufficient to mitigate such risks.

Accepting this principle though, would imply accepting a possible slowdown in the pace of technological innovation. The EU, instead, is struggling to strike a balance between maximum safety for the individual and a democratic society, and minimum interference with the advancement of technological innovation by other means. It will be argued, though, that at least in some cases the anticipatory protection afforded by such principle is unavoidable to prevent the violation of fundamentals rights, and that the recent practice of the Commission opens a window for its application in the assessment of AI systems.[7]

Over time the EU Legislator has engineered a combined regulatory and governance framework aimed at regulating AI technology while monitoring its evolution so to keep the legal framework updated and effective to address possible new risks (this, at least, is the wishful aim of the strategy). The outcome is a complex ecosystem of measures that hinges in the preexisting regulation of the EU internal market, geared towards fostering free movement of goods and services in an environment where both consumer protection and competition among businesses are assured. Of this bundle of measures, the AI Act – Regulation 2024/1689 of June 2024 – represents a major component, but it would be misleading to identify only with this piece of regulation the normative response of the EU to the challenges of AI technology.

The main building blocks of the European framework may be identified in the following: the horizontal protection of personal data provided under the GDPR, that the AI Act intends to strengthen;[8] the horizontal protection of consumers' safety and trust under the General Product Safety Regulation, operationalized by the mechanism for the surveillance of the internal market;[9] the regulation of actors already active in the internal market, whose size entails the capacity to carry a systemic risk (large platforms), provided under the Digital Service Act (DSA).[10]

Instrumental to the overall success of the EU strategy on AI are two additional elements: the empowerment of both Member States' citizens and businesses for a sound uptake of the technology and the commitment to establish the European approach on AI as the prevailing standard globally.

The aim of the present study is to contribute to the understanding of how effective this complex normative framework is in fulfilling its declared objective of shaping a digitalized society where the uptake of AI systems does not endanger the pervasive respect of the fundamental rights and democratic values that, since the end of the second world war, Europe has chosen to put at the center its social and legal order.

After a concise survey of the main features of the EU legal framework on AI (Section 2) and of two constitutive elements for its success, respectively, the creation of a society capable of making a sound

[7] See, *infra*, Section 4.

[8] For instance, through some of the prohibitions in Article 5 of the *AI Act*.

[9] Provided under *Regulation 2019/1020 of the European Parliament and of the Council of 20 June 2019 on market surveillance and compliance of products and amending Directive 2004/42/EC and Regulations (EC) No 765/2008 and (EU) No 305/2011*, *O.J.* L 169, 25.6.2019, 1.

[10] *Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act)*, *O.J.* L 277, 27.10.2022, 1.

use of such disruptive technology at regional (Section 2.1) and at the international level (Section 2.2), the paper will tackle the effectiveness of the claim of consistency of the EU framework with the protection of fundamental rights, democratic values and the rule of law (Section 3). Some final consideration will be offered as a contribution to the discussion on what legal tools might be better suited to shape a rights-oriented society that takes advantage of AI without impairing its values and on the overall capacity of the EU legal order to achieve this result in the implementation of the complex normative framework discussed in the article.

2. The main components of the European AI ecosystem: an integrated normative network

As mentioned above, European AI regulation is entrusted to a network of integrated and complementary regulatory instruments. The AI Act, which is the focus of these brief notes, is part of the body of rules governing the internal market according to the so-called "New Approach". Under this method, which aims to facilitate access to and circulation of products on the internal market, the essential characteristics of products are harmonized in legislative acts (mainly directives), while technical characteristics are defined in standards negotiated by stakeholders within European standardization bodies. It is the direct responsibility of economic operators to verify the conformity of their products with these rules and standards before placing them on the Union market, while national authorities are responsible for *ex post* control and Union authorities have a supervisory role.

In line with this approach, the AI Act qualifies as an instrument designed to regulate access to the European market for AI systems and, accordingly, finds its legal basis in Article 114 of the Treaty on the Functioning of the European Union, as well as, given the centrality of data in the economy of AI systems, in Article 16 of the same treaty.

According to the scheme described here, compliance verification for AI systems is entrusted to economic operators, with obligations divided among them based on the role each plays in the life cycle of the system in question, with *ex post* control and supervision tasks shared between national authorities and EU institutions and bodies.

The protection of individuals and society from the risks connected to AI is sought, first, adopting a risk-based regulatory approach. AI systems are prohibited in situations where they generate risks considered unacceptable (Article 5); they need to comply with strict requirements when they are classified as 'high risk' and are subject to transparency obligations when deemed capable of causing only limited risk (Article. 50). If the risk is minimal, only the voluntary adhesion to codes of conduct is incentivized (Art. 95, AI Act).

General-purpose AI models (GPAIs) are regulated separately and subject to obligations of transparency and compliance with EU copyright legislation.

A separate strategy aims to prevent systemic risks identified in relation to significant impact on the internal market due to negative effects "on public health, safety, public security, fundamental rights, or society as a whole".[11] While the AI Act addresses systemic risks that may be caried by GPAIs by

[11] 'Systemic risk' is defined as "a risk that is specific to the high-impact capabilities of general-purpose AI models, having a significant impact on the Union market due to their reach, or due to actual or reasonably foreseeable

introducing additional constraints (Article 55, AI Act), another regulation, known as the Digital Services Act,[12] aims to govern systemic risks that are particularly dangerous in relation to the size of the economic operators involved.[13]

The effectiveness of the overall framework relies on the deterrence effect of heavy sanctions in case of infringement, but especially on an implementation mechanism which branches out to the Union and Member States level, whose design should ensure persistent conformity with the regulation of the AI systems placed on the internal market, and enable the update of the legislation when needed.

A supervisory and oversight function is entrusted with the AI Office, internal to the Commission (which has a central role for GPAIs and in the enforcement mechanism of compliance of providers of very large online platforms and of very large online search engines under the DSA) and the AI Board at the regional level, and with competent authorities of the Member States at national level. These are called to establish an internal administrative structure branched into bodies responsible for steering the conformity assessment of AI systems (notifying authorities); for market surveillance and for the protection of fundamental rights, while an amendment to the original Commission proposal has introduced a much-welcome procedure of compulsory fundamental rights impact assessment for selected high-risk AI systems.

What needs to be underscored is that this system integrates into the well-established mechanism of surveillance of the internal market, designed to allow both free circulation of non-food products and consumers protection.[14] The last objective (consumer protection) is additionally pursued through the so-called General Product Safety Regulation (GPSR),[15] whose new text, entered into force in December 2024, pays close attention to the regulation of products embedding new technologies and potentially capable of generating unknown risks, including to mental health and cybersecurity of the product.[16] This means that the protection afforded to consumers by the GPSR operates 'as a safety net' for consumers-users of AI systems that, not being qualified as high-risk, are not subject to the conformity procedure provided under the AI Act.[17]

The AI Act is also complemented by the 'Unfair Commercial Practices Directive', whose application may cover subliminal and manipulative practices escaping from the prohibition provided under Article 5(1)(a)

---

negative effects on public health, safety, public security, fundamental rights, or the society as a whole, that can be propagated at scale across the value chain" (Article 3 (65), *AI Act*).

[12] Regulation (EU) 2022/2065.

[13] Providers of very large online platforms and of very large online search engines, as defined in DSA Article 33.

[14] Regulation 2019/1020.

[15] *Regulation (EU) 2023/988 of the European Parliament and of the Council of 10 May 2023 on general product safety, amending Regulation (EU) No 1025/2012 of the European Parliament and of the Council and Directive (EU) 2020/1828 of the European Parliament and the Council, and repealing Directive 2001/95/EC of the European Parliament and of the Council and Council Directive 87/357/EEC*, *O.J.* L 135, 23.5.2023, 1. According to Article 2(1) of the GPSR, "(i)f products are subject to specific safety requirements prescribed by Union law, this Regulation shall apply only to aspects and risks or categories of risks not covered by those requirements", so integrating – horizontally – specific 'vertical' regulations. The interaction of the *AI Act* with Union harmonization legislation is instead regulated under Article 2 (2) of the *AI Act*.

[16] More specifically, the GPSR now covers new risks inherent in "digitally connected products, including mental health, to which consumers are exposed during the provision of a service" (Rec. 23), as well as risks "arising from external interventions affecting the product" (cybersecurity risks, Recs. 24-26 and Article 6.1g).

[17] Cons. (166), *AI Act*.

of the regulation.[18] The interrelation of the AI Act with a network of preexisting legal tools is specifically engineered by the European Legislator, and fundamental in assessing the overall protection offered by the European legal framework against the risks posed by AI systems. This legal architecture is highlighted by the Guidelines on prohibited AI practices, that for each of the cases listed in Article 5(1) clarify which pieces of legislations interact with the AI Act, providing for complementary protection out of the (too strict) scope of the prohibitions.[19]

The complexity of this network of legal tools, that may converge in the regulation of same real-life situations, might need to be rationalized by the legislator, to bring a simplification that does not imply diminished oversight on persons' rights and societal values.

A notable aspect of the EU's AI governance strategy is the extensive powers granted to the Commission. These powers go beyond the usual implementation responsibilities of the institution in the EU legislative process, and are instead aimed at two innovative objectives: ensuring that the regulations are updated in line with technological progress, and encouraging and facilitating compliance with the new discipline.

The first objective is made unavoidable by the choice to regulate a field (the access to the internal market of AI systems) whose future developments cannot be anticipated scientifically.[20]

In addition to recourse to soft law,[21] which is instrumental in addressing both needs, the critical issue of keeping pace with technological advancement has been addressed by empowering the Commission to update the list of use cases for high-risk AI systems in Annex III to the AI Act. This allows new AI systems that pose risks to health and safety or have an adverse impact on fundamental rights to be included when needed (Article 7, AI Act).

In carrying out this role, the Commission is supported by a Scientific panel of independent experts and an Advisory Forum supplying technical expertise. By providing for the presence of these two technical bodies, tasked with keeping the European regulator informed of progress but also of the inherent new risks associated with the use of artificial intelligence systems, the AI Act enables the Union to determine the level of risk deemed acceptable and to update the legislation accordingly. The objective of intervening – by increasing the prohibited cases referred to in Article 5 of the AI Act – before human rights and democratic values are violated by unregulated technological progress, and the expectation that this should be done with the scientific support of technical bodies, corresponds to the precautional

[18] *Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council ('Unfair Commercial Practices Directive'), O.J.* L 149, 11.6.2005, 22.

[19] Communication from the Commission, *Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act)*, C (2025) 5052 final, Brussels, 29/7/2025.

[20] See the International AI Safety Report of January 2025.

[21] On the recourse to soft law, that not needing to undergo lengthy legislative procedure for its adoption and not having necessarily a governmental origin, is particularly fit to address the new risks related to technical innovation, involving private economic operators and stakeholders in the regulatory effort, see M. Sosa Navarro, *The Role of Soft Law in the Regulation and Governance of Human Rights Challenges Posed By Neurotechnology*, Torino, 2025, 37 ff.

logic that has become prevalent in various areas of international law for the protection of the so-called global commons.[22]

Coming to the objective of encouraging and facilitating compliance with the new discipline, this corresponds to a collaborative approach that the Commission is adopting lately to enable European businesses and society at large to take on both the digital and the green transitions. Faced with the new burdens that achieving sustainability and digitization goals impose on businesses, the Commission is stepping in by offering facilitation tools (guidelines, information platforms, uniform models for fulfilling obligations) and direct assistance services to businesses, such as dedicated help desks.[23]

To this end, the AI Act establishes the European Artificial Intelligence Board, whose task is to 'advise and assist the Commission and the Member States to facilitate the consistent and effective application of this Regulation' (Article 65, AI Act), while also incentivising and facilitating compliance through guidelines and codes of conduct. The use of soft law instruments aims to increase and widen the scope of application of the legal requirements under the AI Regulation (encouraging businesses to adhere to codes of conduct).

Furthermore, the Commission can contribute to legal certainty through common specifications (Art. 41) in cases where EU standardisation bodies fail to provide the necessary harmonised standards to streamline implementation and assure conformity.

At the same time, the Union has launched a series of initiatives to boost technological innovation, including the establishment of regulatory sandboxes[24] and support for investments and funding.

2.1. The empowerment of EU society

Another important feature of the EU strategy is its aim to enable European society to profit from tech innovation, while avoiding backlashes. A sound use of digital technology, let alone AI, by the different components of EU society, requires enabling individuals not only to technically handle digital devices and AI systems, but primarily to understand the challenges that a misuse of such technologies can pose to personal rights and freedoms.

This policy goal is pursued by the European Union on two different layers: the first pertains to the generality of EU society, whose AI literacy needs to be increased; the second specifically addresses businesses.

The empowerment of EU citizens to reap the benefits from AI starts with providing them with adequate digital education. In the European Declaration on Digital Rights and Principles for the Digital Decade, the European Parliament, the Council and the Commission jointly committed to reach this goal, stressing the importance to include in the competences of EU learners and teachers the development of "critical thinking".[25] The Declaration makes it clear that such competences are indispensable to empower

[22] This is the case in international environmental law and in the multilateral trading system: for a discussion on this point, see C. RAGNI, *Scienza, diritto e giustizia internazionale*, Milano, 2020, 58 ff.

[23] For instance, the "single Help desk" established to ease compliance with the Directive (EU) 2024/1760 (*Corporate Sustainability Due Diligence Directive* – CSDDD).

[24] *AI Act*, Article 57.

[25] See the signed version of the Declaration. The reference to "critical thinking" was not to be found in the Commission proposal: see Chapter II of the *European Declaration on Digital Rights and Principles for the Digital Decade*, COM (2022)28 final, Brussels, 26/1/2022. For a comment on the Declaration, see: A. ADINOLFI, *Evoluzione*

individuals to participate in the digital public space making "freely given, specific choices" and that peoples' empowerment encompasses the ability to use algorithms and AI and to be informed when interacting with them and to acquire control on how their personal data are used and with whom they are shared.[26] Specific initiatives to foster an advanced digital education and to increase the skills of the citizens in managing digital products and services are set up to reach these outcomes.[27]

It is the AI Act directly, instead, that provides for fostering AI literacy in the business environment, through its Article 4, that entrusts providers and deployers with this task in respect of their "staff and other persons dealing with the operation and use of AI systems on their behalf".

In pursuing the competent involvement of society in the so called "digital transformation", the EU relies on the tools introduced by the Interinstitutional Agreement on better legislation, starting from public consultations at different levels (targeted at experts or open to the public)[28] and structured dialogues with industry.

This participatory 'whole of society approach' has been recently revised by the Commission through the *AI Continent Action Plan*.[29] The Communication outlines initiatives aimed at creating an ecosystem that encourages the uptake of the technology by businesses facilitating access to the necessary infrastructures.[30]

Initiatives include a *Data Union Strategy*, which aims to strengthen "interoperability and data availability across sectors, to respond to the scarcity of robust and high-quality data for the training and validation of AI models", as well as the establishment of *Data Labs,* where high-quality data from *AI Data Factories* related to the same sector could be 'federated' and linked to corresponding *EU Data Spaces*, allowing developers, under certain conditions, to access large amounts of them. The idea is to create public environments where technological innovation can be driven by secure access to the necessary data. Adoption of the regulation establishing the European Health Data Space has signalled the start of implementation of this project.[31]

2.2. The promotion of a shared approach to AI at international level

In accordance with Article 21 of the Treaty on European Union, the EU pursues consistency between its internal and external actions, including those relating to the development and regulation of AI. This

tecnologica e tutela dei diritti fondamentali: qualche considerazione sulle attuali strategie normative dell'Unione, in *I Post di AISDUE*, V, *Quaderni AISDUE*, 2023, 321-330.

[26] See the final text of the Declaration cited in the previous note, at Chapter IV, III and V respectively.

[27] See the Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions Digital Education Action Plan 2021-2027. *Resetting education and training for the digital age*, COM (2020) 624 final, Brussels, 30/9/2020.

[28] E.g., the Commission opened a targeted consultation seeking input on guidelines to clarify rules for general-purpose artificial intelligence (GPAI) models under the EU *AI Act*.

[29] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions *AI Continent Action Plan*, COM (2025) 165 final, Brussels, 9/4/2025, 15.

[30] The EU strategy encompass the allocation of funds, that should elicit private investment.

[31] *Regulation (EU) 2025/327 of the European Parliament and the Council of 11 February 2025 on the European Health Data Space and amending Directive 2011/24/EU and Regulation (EU) 2024/2847*, *O.J.* L series, 5/3/2025.

means that the EU's objective of promoting the use of human-centered, trustworthy AI systems is reflected in its relations with third countries and its actions in international *fora*.

Furthermore, fostering a shared vision of AI that mitigates risks stemming from divergent perspectives on its role in society has become of paramount importance in light of the current geopolitical tensions. This is why nurturing cooperation with third countries and organisations to converge on shared principles of AI regulation at an international level has become so important.

To achieve this goal, the Union is pursuing two parallel lines of action: cultivating collaborative relationships to support adherence to shared principles, technological advancement and the interoperability of respective digital infrastructures at the bilateral level; actively participating at the international level in multilateral initiatives to support the establishment of a human-centric and trustworthy model of AI.[32]

The first type of actions can be traced back to the network of digital partnerships and alliances with partner countries across the world, formalized through Ministerial-level Trade and Technology Councils,[33] Digital Partnerships, Digital and Cyber Dialogues, or specific chapters on digital trade in more comprehensive trade and association agreements.[34] Important aspects of the Union's digital diplomacy at bilateral and regional level relate to strengthening cybersecurity and cooperation in the field of security and defense, and promoting European models within international standardization bodies.[35]

At international level, the EU participates in all the main *fora* promoting a multilateral and multi-stakeholder approach in addressing the challenges of technological innovation. While it is beyond the scope of this paper to analyze the position taken by the Union at the various tables, it seems appropriate to briefly mention the two main ones, the Council of Europe at the regional level and the United Nations.

With regard to the former, it is important to note the complementary relationship between the AI Act and the Framework Convention on Artificial Intelligence adopted by the Council of Europe in September 2024.[36]

The result of parallel and mutually influential legislative processes,[37] the two regulatory instruments appear to complement each other in such a way as to compensate for their respective weaknesses. As it

[32] At the request of the European Council, the Commission and the High Representative of the Union for Foreign Affairs and Security Policy reported on this dual approach in the Joint Communication to the European Parliament and the Council *An International Digital Strategy for the European Union*, JOIN (2025) 140 final, Brussels, 5/6/2025

[33] It is worth mentioning that the EU-U.S. Trade and Technology Council (TTC), in order to better align the approaches to risk management and trustworthy AI of the two countries and to advance their collaboration in international standards bodies has been able to express a *EU-U.S. Terminology and Taxonomy for Artificial Intelligence*, first adopted in May 2023 and reviewed through a participatory process in 2024.

[34] For an overview of the existing agreements, see the Joint Communication, *op. cit.*

[35] On the relevance of standardization in achieving EU political goals: P. CHION, *Standards for AI Governance: International Standards to Enable Global Coordination* in *AI Research & Development*, Future of Humanity Institute. University of Oxford, 2019, retrievable from https://cdn.governance.ai/Standards_-FHI-Technical-Report.pdf.

[36] Council of Europe *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*, CETS 225, Vilnius, 5/9/2024, retrievable from https://rm.coe.int/1680afae3c.

[37] For an in-depth analysis of the process of mutual influence between the Council of Europe and the EU, and for the argument that, with regard to AI regulation, it is the EU instrument that ensures the most effective protection of fundamental rights, see F.P. LEVANTINO, F. PAOLUCCI, *Advancing the Protection of Fundamental Rights through AI*

is natural, given the different nature of the issuing organizations,[38] the European regulation has 'more teeth' with regard to market operators, who are directly subject to obligations and controls. The convention, for its part, is more rigorous in reiterating the obligation of member States to ensure respect for human rights and democratic principles in the use of AI in activities "related to the protection of its national security interests" that fall outside the scope of the treaty.[39]

The United Nations has spoken with multiple voices on the subject of artificial intelligence. UNESCO outlined in a Recommendation the principles that must inspire Members when addressing such technology.[40] The Secretary-General launched a process to institutionalize the Organization's action in the field of technological innovation, which led to the establishment of the Office for Digital and Emerging Technologies and the formulation of new recommendations in the report Governing AI for Humanity.[41] The General Assembly, as part of the Summit for the Future in September 2024, promoted the adoption of the Global Digital Compact.[42]

Overall, the EU's policy on AI appears to align with the recommendations expressed by the UN. However, regarding the need to prevent unequal access to AI technology from exacerbating the digital divide with the Global South, which is emphasised in UN documents, the EU's recently disclosed international strategy seems particularly weak.

3. The claim to a 'human-centered' regulation of AI: a critical assessment

Notwithstanding the impressive normative effort summarized in the previous Sections, when we analyze the current EU framework on AI through the lens of human rights, in order to assess le level of protection that such framework can grant to the dignity and autonomy of the person, identified by the Council of Europe as the core of humaneness,[43] there is more than one reason of concern. Many of the regulatory instruments through which the AI Act aims to achieve the objective set out in Article 1 appear, in fact, to be ineffective for this purpose.

Regulation: How the EU and the Council of Europe Are Shaping the Future, in: *European Yearbook on Human Rights* 2024, 3-37.

[38] The different scope and potential effects of the two instruments due to their legal nature is analyzed by J. ZILLER, *The Council of Europe Framework Convention on Artificial Intelligence vs. the EU Regulation: two quite different legal instruments*, in *CERIDAP*, *Rivista Interdisciplinare sul Diritto delle Amministrazioni Pubbliche*, 2/2024, 202.

[39] Framework Convention, Article 3(2).

[40] United Nations Educational Scientific and Cultural Organization (UNESCO), *Recommendation on the Ethics of Artificial Intelligence*, SHS/BIO/REC-AIETHICS/2021, 12/11/2021, available at https://unesdoc.unesco.org/ark:/48223/pf0000380455.

[41] The document was issued by the High-level Advisory Body on Artificial Intelligence in September 2024. The potential negative impact of AI systems on human rights is insightfully analyzed by A. KRIEBITZ, C. LUTGE, *Artificial intelligence and human rights: business ethical assessment*, in *Business and Human Rights Journal*, 5(1), 2020, 84 ff.

[42] The text of the Compact can be accessed at https://www.un.org/global–digital–compact/sites/default/files/2024–09/Global%20Digital%20Compact%20–%20English_0.pdf.

[43] See Article 7 of the *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*.

The decision to restrict the compulsory fundamental rights impact assessment (FRIA) under Article 27 to only a small fraction of high-risk AI systems is not justified.[44] It has been noted, though, that not only AI systems classified as AI risk may threaten fundamental rights and should be subject to an assessment procedure under the AI Act.[45]

Furthermore, the entire assessment process, as outlined in Article 27 (obligation of the deployer to submit a filled-out template, set to "facilitate deployers in compliance"), resembles a mere formality.

The weakness of the FRIA is particularly concerning given that the general oversight of Member State authorities protecting fundamental rights, as set out in Article 77, is at risk of being ineffective. This is not only due to the fact that several countries are lagging in implementing the norm and have missed the November 2024 deadline to identify and notify their list of authorities to the Commission. More importantly, however, the discretion given to EU Members to choose between jurisdictional and administrative control seriously undermines the effectiveness of the FRIA.

The overall architecture outlined to ensure that AI systems introduced to the European internal market comply with the AI Act relies too heavily on self-assessment by economic operators in the absence of credible public-interest oversight.[46]

While the Commission's implementing powers may be used to address the reported weaknesses, the pressure on EU institutions not to hinder the growth of the European digital economy seems to be leading the Commission towards a too-lenient interpretation of the prohibitions in Art. 5 of the AI Act. It has already been noted that, due to the numerous exceptions that limit their scope, the prohibitions listed in Art. 5 of the AI Act can be considered mere restrictions.[47]

In light of the seriousness of the rationale behind the interdiction of the eight AI practices listed in Article 5 of the AI Act, as set out in Recital 28 of the Regulation, and given the reduction in their scope determined by the limitations set out in Article 5 itself, the Commission's assertion that its recently issued guidelines "strive to interpret the prohibitions *in a proportionate manner* that achieves the objectives of the AI Act to protect fundamental rights"[48] seems to pave the way for an overly restrictive interpretation of the prohibited practices.

---

[44] More specifically, those deployed by bodies governed by public law, private entities providing public services, or those intended to be used to evaluate the creditworthiness of private persons or for risk assessment and pricing in relation to natural persons in the case of life or health insurance (Article 27 (1) and Annex III (5) (b),(c)).

[45] LONGO and PAOLUCCI argue that a comprehensive FRIA is necessary to mitigate the human rights risks posed by deepfake technologies, which cannot be adequately addressed by the assessment procedures set out in the GDPR and DSA: E. LONGO, F. PAOLUCCI, *The Article 50 of the AI Act and the Transparency Obligations: The Model and its Limitations*, in O. POLLICINO, F. DONATI, G. FINOCCHIARO, F. PAOLUCCI, *La disciplina dell'intelligenza artificiale*, cit., 275-295.

[46] POLLICINO observes that FRIA, by limiting itself to requiring notification by the deployer to the Market Surveillance Authority, does not provide for external control and thus reproduces the structural shortcomings of the other impact assessments referred to in the GDPR and the DSA: O. POLLICINO, *Regolare l'intelligenza artificiale: la lunga via dei diritti fondamentali*, in O. POLLICINO, F. DONATI, G. FINOCCHIARO, F. PAOLUCCI, *La disciplina dell'intelligenza artificiale*, cit., 3-35.

[47] F.P. LEVANTINO, I. NERONI REZENDE, *Rischio inaccettabile*, in O. POLLICINO, F. DONATI, G. FINOCCHIARO, F. PAOLUCCI, *La disciplina dell'intelligenza artificiale,* cit., 159-160.

[48] Communication from the Commission, *Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act)*, C(2025) 5052 final, Brussels, 29/7/2025, point (93).

This concern is confirmed, for instance, by the given interpretation of the concept of 'significance' in relation to the physical harm that may be caused by manipulative practices.[49] According to the guidelines, a "significant physical harm that is reasonably likely to be caused by an AI system" occurs when

> AI systems […] suggest to an individual to commit criminal acts such as sexual abuse and exploitation, extreme violent or terrorist content or incentivize individuals to commit crimes, self-harm or harm to other persons […]. By contrast, minor physical harms may include less severe injuries, such as bruises or temporary discomfort, which do not have significant or lasting consequences and will therefore not reach the threshold of significance within the meaning of Article 5(1)(a) AI Act.

It is at least questionable that this interpretation respects the right to the integrity of the person enshrined in Article 3 of the EU Charter of Fundamental rights. Rather, the Commission should have made clear that any physical or psychological harm would be 'significant', limiting the category of 'un-significance' to economic/financial harm.

4. Conclusions: *plaidoyer* for a strengthened precautionary approach

The complex framework set up by the European Union to channel the development of AI technology into paths that are values and rights oriented still needs to be fine-tuned, if it wants to keep its promises.

While the declared 'human-centric' approach is highly embraceable, its implementation is not equally convincing.

The main cause for concern is the market-driven premise that the uptake of AI must be incentivised in every area of life in order to determine an intrinsic transformation of society, or the 'digital transformation'. In other words, it is not considered to be – as it should – a powerful tool that can help individuals, researchers and businesses achieve their respective goals more easily and with better results, thereby contributing to societal progress. Instead, it is conceived as a disruptive innovation that is doomed to change individuals' lives and society in ways that might not be reconcilable with their long-cherished values and aims.

More specifically, concerns can be raised about the potential consequences for our society and the environment of the political decision to promote the widespread deployment of AI systems in the European market, which has led the legislator to limit only to scanty cases the assessment of their impact on fundamental rights and democratic principles.

As EU regulation stands now, individuals are not satisfactorily protected from the uptake of AI in education that might endanger the cognitive and social development of children,[50] nor from AI systems

[49] "(S)ubliminal techniques beyond a person's consciousness or purposefully manipulative or deceptive techniques, with the objective, or the effect of materially distorting the behaviour of a person or a group of persons by appreciably impairing their ability to make an informed decision, thereby causing them to take a decision that they would not have otherwise taken in a manner that causes or is reasonably likely to cause that person, another person or group of persons significant harm" (Article 5(1)(a) of the *AI Act*).

[50] Besides the benefits, possible negative impacts of AI on children's cognitive, social, and emotional development, and mental health are analyzed by G. Osório de Barros, O. Severino Soares, *AI and the Next Generation: Protecting*

replacing workers in businesses, which could disrupt the European social model based on national social security systems. The planned massive increase in AI farms in EU countries, with their huge consumption of water, risks to clash with the goals pursued with the EU Green Deal and more generally with States' obligations to prevent significant harm to the environment and take the lead in combating climate change.[51] The fact that AI can be successfully deployed for aims of environmental protection (e.g. for analyzing the soil) is not a valid counter-argument: given the heavy weight of the technology on the consumption of water resources, a sound response of the legal system would be the introduction of clear boundaries for its use.[52]

The exclusion of the military and defense sectors from the scope of the AI Act constitutes a major *vulnus* to the plausibility of the human-centered nature of the framework that only a serious effort by the Union to enter into meaningful negotiations at international level to fill this regulatory gap might mend.[53]

The normative framework intended to afford protection of fundamental rights and societal values like democracy and the rule of law seems, at the same time, too cumbersome and too inefficient.[54] Still, much can be done in the implementation phase to fix inconsistencies and strengthen the overall capacity of the EU framework to foster an uptake of the AI technology in European society coherent with those rights and values.

The first step would be to offset the overreliance on self-evaluation by economic operators in the AI Act by strengthening oversight by national authorities. To this end, the Commission could use its substantial enforcement powers to establish common benchmarks following consultations with human rights experts and ethicists. These benchmarks would then be applied by national authorities. Such intervention by the Commission would also reduce the imbalance in human rights protection between Member States that have appointed administrative or jurisdictional national authorities, thereby ensuring consistency in the objective content of the assessment.[55]

The central role that the Commission can play in promoting an application of the AI Act that is more focused on ensuring the protection of fundamental rights can be perceived in the recently published Guidelines on prohibited artificial intelligence practices. Here, in interpreting the concept of "significant harm" on which the prohibition is based, both in relation to the use of subliminal and deceptive techniques likely to compromise the ability of a person or group to make a decision intentionally (Article

---

Childhood in the Digital Age, in A.D.B. MACHADO *et al.* (Eds.), *Environmental, Social, Governance and Digital Transformation in Organizations, Information Systems Engineering and Management*, 35, 2025.

[51] Such obligations have been recently affirmed by the International Court of Justice (ICJ) in the Advisory Opinion released on the request of the United Nations General Assembly on the 23 July 2025: ICJ, *Obligations of States in Respect of Climate Change*, Advisory Opinion of 23 July 2025, accessible at https://icj-web.leman.un-icc.cloud/sites/default/files/case-related/187/187-20250723-adv-01-00-en.pdf.

[52] It might be limited, for example, the use of AI in the field of entertainment, where apps enabling to easily make fake videos with unaware persons acting in situations that might even imply their legal responsibility seem to be headed for success: https://www.nytimes.com/2025/10/09/world/artificial-intelligence-slop.html.

[53] O. POLLICINO, *Regolare l'intelligenza artificiale: la lunga via dei diritti fondamentali*, cit., 34.

[54] See, for a discussion of this point, the previous Section.

[55] The argument is proposed by F. PAOLUCCI, *The Enforcement of the Artificial Intelligence Act: Looking Forward to a Commission Implementing Decisions for Protecting Fundamental Rights*, in O. POLLICINO, F. DONATI, G. FINOCCHIARO, F. PAOLUCCI, *La disciplina dell'intelligenza artificiale*, cit., 341-373.

5(1)(a)), and referring to practices aimed at changing the behavior of an individual or group by exploiting their vulnerabilities (Article 5(1)(b)), the Commission refers for the first time, in relation to AI, to the precautionary principle.[56]

It is significant that the Commission expressly refers to Article 191(2) of the Treaty on the Functioning of the European Union, which codified the precautionary principle in environmental policy. According to the Commission, reading the objective of the AI Act to ensure 'a high level of protection', in conjunction with Article 191(2) TFEU, "suggests a comprehensive approach to protection when assessing the significance of the harm. This means considering both immediate and direct harms and systemic, indirect adverse impacts associated with AI systems deploying subliminal, purposefully manipulative or deceptive techniques that are intended to or capable of impairing individual autonomy, decision-making and free choices of persons and groups of persons." (Guidelines, point (93)).

The unexpected recourse by the Commission to the precautionary principle is promising. The seriousness of the impact on people's fundamental rights and the consequences for the democratic functioning of the State that the abuse or malfunctioning of AI systems can cause requires preventive action.

The *ex post* protection deriving from the possibility for judges to directly apply the provisions of the European Charter of Fundamental Rights, recently reaffirmed by the Court of Justice,[57] in the interpretation of the AI Act would not be effective. Hence the importance of market operators themselves and the authorities vested with powers to supervise their activities adopting a precautionary approach in the preventive assessment of the compliance of AI systems with the Union's system of rights and values. Moreover, the precautionary principle has established itself internationally as an indispensable tool for protecting global commons such as health, the environment and the climate in the absence of scientific certainty.

In the EU legal system, too, its scope has expanded beyond the environmental sector to encompass health protection more broadly.[58] In practice, applying the precautionary principle may mean that, when faced with the risk that new technologies could seriously undermine fundamental rights, the only option is to refrain from using them until it is scientifically certain that they can be used safely.[59] Although the principle has been given a context-specific application in different areas of EU law, a review of EU case law shows that, while purely hypothetical risks are irrelevant, the sufficiency of scientific uncertainty

[56] *Commission Guidelines on prohibited artificial intelligence practices*, cit.

[57] Corte di Giustizia, 3 giugno 2025, in causa C-460/23, *Kinsa*, 68-72.

[58] *Communication from the Commission on the precautionary principle*, COM(2000) 1 final, Brussels, 2/2/2000. The evolution of the precautionary principle from a philosophical concept to a real normative institute of the European legal order was affirmed by the European Court of Human Rights – ECtHR in *Tătar v. Romania,* Application No. 67021/01, 27 January 2009) at para. 69: "l'évolution du principe d'une conception philosophique vers une norme juridique". For an analysis of the ECtHR case law dealing with the tension between human rights and technological innovation, see: T. Murphy, G. O Cuinn, *Works in Progress: New Technologies and the European Court of Human Rights,* in *Human Rights Law Review* 10, 4, 2010, 601 ff.

[59] C. Bublitz, J.A. Chandler, F. Molnár-Gábor, M. Sosa Navarro, P. Kellmeyer, S.R. Soekadar, *A Moratorium on Implantable Non-Medical Neurotech Until Effects on the Mind are Properly Understood*, in *Neuroethics*, 2025.

may vary according to the value or right at risk of negative impact.[60] The call for caution raised by leading members of the scientific community regarding the potential negative impact of high-risk AI systems on individuals' rights, well-being, and democratic infrastructures suggests that the precautionary principle is the most appropriate means of reconciling these technologies with the protection of human rights.

In the technological context, beyond its explicit formulation, the precautionary principle is implemented through a procedure that the OECD calls "anticipatory governance", which involves clearly defining guiding values and the conduct, with the participation of stakeholders, of "strategic intelligence" based on "(r)obust tools such as horizon scanning, advanced data analytics, forecasting, and technology assessment (that) should be employed to anticipate future challenges and inform governance strategies."[61] The roots of such an approach are already embedded in the regulatory fabric of the Union (FRIA, regulatory sandboxes, empowerment of EU citizens and businesses allowing them to act as meaningful stakeholders) and now require to be broadened in their scope and rigorously implemented.[62]

Lastly, the absence of EU-level rules on who would be responsible for negative impacts of the technology risks undermining the system's credibility and trustworthiness, while also damaging businesses due to legal uncertainty and difficulty operating within 27 different legal systems with different liability rules. Providing uniform rules on the liability of AI system operators, in addition to the cases of contractual liability that are already regulated, would undoubtedly improve enforcement effectiveness.[63]

In conclusion, the tension between AI development and the guarantee of rights remains unresolved, requiring constant critical and constructive oversight from civil society and academia.

[60] A comprehensive analysis of the evolution of the precautionary principle in EU law and case law is conducted by K. DE SMEDT, E. VOS, *The Application of the Precautionary Principle in the EU,* in H.A. MIEG (Ed.), *The Responsibility of Science*, Berlin, 2022, 163 ff.

[61] OECD, *Framework for Anticipatory Governance of Emerging Technologies*, OECD Science, Technology and Industry Policy Papers, OECD Publishing, Paris.

[62] Civil society organisations have called on the Commission to take action to require Member States that AI governance structures are well-resourced and officially designated, and that civil society is actively engaged in and embedded within them: EDRI, *Open Letter: The European Commission and Member States must keep AI Act national implementation on track*, https://edri.org/our-work/open-letter-european-commission-member-states-keep-ai-act-national-implementation-on-track/.

[63] K. ZENNER, *An AI Liability Regulation would complete the EU's AI strategy*, in *CEPS*, 2025. On the applicability of the Digital Content and Services Directive (DCSD) 2019/770 and the Sale of Goods Directive (SGD) 2019/771 to AI systems, see M. EBERS, *Liability For Artificial Intelligence and EU Consumer Law*, in *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*, 204, 2021.

Essays

Regulatory Tipping Point: Key Lessons from a Divergence in AI Regulation Between the EU and the US

Jordan L. Fischer*

ABSTRACT: In the last five years, artificial intelligence technologies have exploded onto the global stage. At the same time, different regions have drastically diverged in the approach to regulating artificial intelligence. This article explores two dominate players in the artificial intelligence world: the European Union and the United States, and discusses, in turn, how each has approached regulating artificial intelligence and lessons learned for the next wave of artificial intelligence development and regulation.

KEYWORDS: artificial intelligence; United States regulation; European regulation; comparative analysis; emerging technologies

SUMMARY: 1. Introduction – 2. Overview of the EU and the U.S. Approaches to Artificial Intelligence Regulation – 2.1. The European Proactive Approach in Adopting the EU AI Act – 2.2. The U.S. Disjointed Approach to AI Regulation – 2.2.1. The U.S. Federal Artificial Intelligence Initiatives – 2.2.2. Existing U.S. Federal Regulations that Impact Certain AI Use Cases – 2.2.3. The NIST AI Risk Management Framework as a Non-Regulatory Response – 2.2.4. The U.S. States Attempt to Create Artificial Intelligence Regulatory Initiatives – 3. A Comparison of the Regulatory Approaches to Artificial Intelligence in the EU and the U.S. – 4. How to Think About How to Regulate Artificial Intelligence – 4.1. Lesson Number One: Existing Laws Do Provide Some Regulatory Protections, Without Even Mentioning Artificial Intelligence – 4.2. Lesson Number Two: Technical Standards and Controls May Offer a Middle Ground, With Flexibility and Ease of Adoption – 4.3. Lesson Number Three: The Development of Artificial Intelligence Is Isolated to Only a Few Regions in the World – 5. Conclusion.

1. Introduction

In the last five years, the use of artificial intelligence exploded across the globe, impacting individuals in their daily lives, businesses in their day-to-day operations, and dominating discussions at all levels. For the most part, this technology remained, and continues to remain, unchecked, with major players innovating without guardrails or regulatory requirements.

Within this backdrop, the European Union (EU) adopted the EU Artificial Intelligence Act (the EU AI Act), the first attempt to create a comprehensive regulatory framework around the use of artificial intelligence. However, unlike the EU's adoption of the General Data Protection Regulation in 2018, which became a dominant global force with many countries adopting a similar law in their country, widescale adoption of the EU AI Act has not been embraced, and there is even push back on the EU AI Act within the EU and beyond. On the contrary, many countries are intentionally pausing artificial

* *Drexel University, Thomas R. Kline School of Law. Mail: jlf324@drexel.edu.* *This article was subject to a blind peer review process.*

intelligence regulation, or, as in the United States, intentionally restricting the creation of artificial intelligence laws and regulations.

This article lays out that artificial intelligence regulation is at a unique tipping point: will countries follow a more European model and adopt some regulatory guardrails around the creation and use of artificial intelligence? Or will countries instead allow artificial intelligence to develop unchecked, putting innovation and economic concerns above any potential negative impacts of the rapid adoption of artificial intelligence?

Part I explores the two different, and dominant, regulatory approaches to artificial intelligence: first, the more regulatory heavy focus in the EU and second, the more hands-off approach to artificial intelligence regulation in the U.S. To date, the EU has taken a proactive regulatory approach to the creation and adoption of artificial intelligence. Conversely, the U.S. appears to be focusing solely on innovation and encouraging the development of artificial intelligence, with little to no regulatory guardrails. Part II looks at the comparison of the EU and the U.S. Part III will discuss lessons learned to date, and ways to consider a path forward in regulating artificial intelligence on the global stage.

2. Overview of the EU and the U.S. Approaches to Artificial Intelligence Regulation

It should come as no surprise that the EU and the U.S. take different approaches to regulation, and are diverging dramatically in the context of artificial intelligence.[1] Historically, in Europe, both at the EU level as well as at the Member State level, governments have favored adopting regulation across a wide sector of industries and areas of society. The EU, in some ways, takes a more paternalistic approach, protecting individuals against corporations and the government alike. This is best exemplified by recent regulatory approaches to technology within the EU, creating comprehensive and protective regulations that focus on the individual rights, and protection from corporate actors.[2]

Contrast the EU approach with the U.S., where the focus is more market driven, with data, and subsequently privacy, made into more of a commercial asset as opposed to a regulated right. The U.S. approach to the digital economy is often an "uncompromised faith in markets and skepticism toward government regulation".[3] The U.S. comes from a more techno-libertarian ethos, an approach that emphasizes individual freedom, minimizes government intervention, and focuses on the potential for technology to create a "free" (meaning no regulation, not no cost) and unregulated online environment.

This dichotomy in the approaches between the EU and the U.S. is no more relevant than in the context of the technology industry. There is this widely accepted view in the U.S. that regulation will hamper innovation, a view that is heavily promoted by technology companies themselves. Instead, the U.S. has placed a large reliance on the market and self-regulation to keep technology companies in check. As Anu Bradford sums up, "most governments have refrained from regulating the tech industry precisely

[1] A. ENGLER, *The EU and U.S. diverge on AI regulation: A transatlantic comparison and steps to alignment*, in *Brookings Institution*, 2023, available at https://www.brookings.edu/articles/the-eu-and-us-diverge-on-ai-regulation-a-transatlantic-comparison-and-steps-to-alignment/ (last visited 30/09/2025).

[2] NEWMAN, L. ABRAHAM, *Protectors of Privacy*, in *Cornell University Press*, 2008 (describing the European comprehensive approach to privacy regulation).

[3] A. BRADFORD, *The False Choice Between Digital Regulation and Innovation*, in *Nw.U.L.Rev*, 119, 2024, 377-387.

because of their fear that attempts to interfere with tech companies' operations would undermine their innovative capacity".[4]

This article will explore these divergent approaches to technology regulation (or lack of regulation) in the context of artificial intelligence. The EU, aligning with its historical approach, has taken a regulation first approach to artificial intelligence with its adoption of the EU AI Act. The U.S., conversely, has been slow to adopt regulation for artificial intelligence in the federal context, even briefly considering (and the ultimately rejecting) a formal federal moratorium on artificial regulation for a period of ten years. The lack of federal legislation on artificial intelligence has given way to a state-by-state patch work approach to artificial intelligence regulation. We will explore each in turn.

2.1. The European Proactive Approach in Adopting the EU AI Act

The EU AI Act is touted as the "first regulation on artificial intelligence".[5] The European Commission initially proposed the original draft of the EU AI Act in 2021, incredibly early in the adoption of artificial intelligence technologies across the globe.

It is interesting to place the adoption of the EU AI Act in the context of the artificial intelligence commercial marketplace. ChatGPT first became available to the general public as a free research preview on November 22, 2022.[6] Within two months of its initial release, ChatGPT was estimated to have reached 100 million users, far surpassing the adoption rate for other technologies like Facebook, Instagram, and TikTok.[7] In one year, the number of users grew to more than 100 million, and in two years, that number grew to 350 million users.[8] It is estimated that approximately 10% of the global adult population is using ChatGPT by mid-2025.[9]

The EU AI Act was ahead of the curve, introduced a year before ChatGPT became available to the mass market. While artificial intelligence was already used in certain commercial products,[10] its mass adoption rates coincided with the same timeline for final adoption of the EU AI Act, which was fully adopted in June 2024.

The EU AI Act lays out a set of risk-based rules for the creation and use of artificial intelligence in Europe. It is part of the EU's larger digital strategy to create a comprehensive regulatory approach

[4] BRADFORD, *op. cit.*, 379.

[5] European Parliament, *EU AI Act: first regulation on artificial intelligence*, February 19, 2025, available at https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence (last visited 30/09/2025).

[6] *History Of ChatGPT: A Timeline Of The Meteoric Rise Of Generative AI Chatbots*, in *Search Engine Journal*, 2024, available at https://www.searchenginejournal.com/history-of-chatgpt-timeline/488370/ (last visited 30/09/2025).

[7] *ChatGPT sets record for fastest-growing user base - analyst note*, in *Reuters*, 2023, available at https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/ (last visited 30/09/2025).

[8] A. CHATTERJI *et al, How People Use Chatgpt*, in *NBER*, Working Paper No. 34255, 2025, available at http://www.nber.org/papers/w34255 (last visited 30/09/2025).

[9] *Ivi,* 10.

[10] *What is the history of artificial intelligence (AI)?,* in *Tableau*, available at https://www.tableau.com/data-insights/ai/history (last visited 30/09/2025).

across Europe for a variety of digital areas.[11] This strategy includes the Digital Services Act (DSA),[12] the Digital Markets Act (DMA),[13] and the Data Governance Act (DGA),[14] in addition to the EU AI Act. The EU's General Data Protection Regulation (GDPR),[15] adopted in 2018, was the first regulation in the EU's concerted effort to develop controls around the use of technology, and the outsized role that large technology companies play in everyday society.

Each of these laws are intended to focus on different aspects of the digital space. The DSA attempts to create a safer and fairer digital space by regulating online intermediaries (platforms, marketplaces, social networks, etc.) and enhancing accountability. The DMA focuses on competition in the digital space, and specifically anti-competitive practices by larger technology companies. And, the DGA attempts to create a framework for the sharing and reuse of data, particularly public sector data, while addressing data protection and ethical considerations. The GDPR focuses on the use of personal information across any platform or use case, whether it includes technology or not.

The EU AI Act builds on these other regulations with a specific focus on artificial intelligence technologies. Specifically, the EU AI Act lays out expectations for providers and deployers of "AI systems"[16] to limit the risks of artificial intelligence, especially high-risk use-cases. A provider is defined as 'a natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge'.[17] A deployer is defined as "a natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity".[18]

---

[11] European Commission Digital Strategy, 2022, available at https://commission.europa.eu/publications/european-commission-digital-strategy_en (last visited 30/09/2025).

[12] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (*Digital Services Act*), available at http://data.europa.eu/eli/reg/2022/2065/oj.

[13] Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (*Digital Markets Act*), available at http://data.europa.eu/eli/reg/2022/1925/oj.

[14] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (*Data Governance Act*), available at http://data.europa.eu/eli/reg/2022/868/oj.

[15] Regulation (EU) 2016/679 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, and repealing Directive 95/46/EC (*General Data Protection Regulation*), available at http://data.europa.eu/eli/reg/2016/679/2016-05-04.

[16] An "AI system" is defined as "a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments." EU *AI Act*, Article 3(1).

[17] EU *AI Act*, Article 3(3).

[18] EU *AI Act*, Article 3(4).

There are four risk levels under the EU AI Act: unacceptable risk, high risk, limited risk, and minimal risk.[19] The EU AI Act provides guidance in determining where a given AI tool or use-case may sit within these risk levels, and then provides corresponding requirements based on the risk-level identified.

The ultimate goal of the EU AI Act is two-fold: first, to create "trustworthy AI"; and second, to create governance around the development and deployment of artificial intelligence within the EU. Specifically, the EU AI Act is intended "to promote the uptake of human centric and trustworthy artificial intelligence (AI) while ensuring a high level of protection of health, safety, fundamental rights as enshrined in the Charter of Fundamental Rights of the European Union (the 'Charter'), including democracy, the rule of law and environmental protection, to protect against the harmful effects of AI systems in the Union, and to support innovation".[20]

In many ways, the EU attempted to right a perceived error on their part with the early adoption of the EU AI Act to try to halt the unrestrained development and then mass use of artificial intelligence, as opposed to retroactively apply regulatory controls, as it did with personal information under the GDPR. The GDPR, from a very practical sense, was arguably too late to really restrain the massive data collection, and use, of personal information by private companies, making it challenging to actually enforce privacy protections with these companies that survive on massive data ingestion. Learning from that experience, the EU is attempting to set the guidelines for AI before it becomes a dominate force in society, and almost an inevitable technology we all must live with.

However, the EU AI Act faces numerous hurdles to its enforcement. On the eve of its effective date, numerous stakeholders continue to push for it to be delayed and/or amended before it goes into full effect. For example, in an open letter, forty-five (45) of Europe's largest companies called on the European Commission to pause the EU AI Act's most stringent requirements for two years.[21] Criticism of the EU Act has taken various forms, including a lack of certainty and guidance from the EU on how the EU AI Act will be enforced, the cost of compliance creating huge hurdles for businesses, especially small to medium sized businesses, and the lack of clear standards for businesses to use to create effective compliance to comply with the EU AI Act. The ultimate success of the EU AI Act in creating effective controls on artificial intelligence will play out over the coming years.

2.2. The U.S. Disjointed Approach to AI Regulation

Artificial intelligence is facing a similar path as most technology regulation in the U.S.: a divide between the federal and the state governments. Similar to data protection regulation, which has seen more movement in at the state level than the federal level, the U.S. states have taken a more proactive stance on artificial intelligence regulation as compared to the federal level. However, the federal government has not remained silent on artificial intelligence, providing guidance in other forms outside of traditional regulation.

---

[19] *High-level summary of the AI Act*, in *EU AI Act*, 2024, *available at* https://artificialintelligenceact.eu/high-level-summary/ (last visited 30/09/2025).

[20] *EU AI Act*, Preamble.

[21] *EU AI Champions Open Letter Stop-the-clock to reset the EU's AI ambitions*, July 3, 2025, available at https://docs.google.com/document/d/16570SgWppeeeYINe4WydTbG2ioBTPdOW_fmfMbmg6sY/edit?pli=1&tab=t.nwpblityhtt3.

2.2.1. The U.S. Federal Artificial Intelligence Initiatives

At the federal level, the U.S. government approach to artificial intelligence has evolved under the two most recent presidential administrations. While the Biden administration focused on creating certain guardrails to the development and use of artificial intelligence, the most recent Trump administration is promoting innovation of artificial intelligence technologies over the development of any artificial intelligence guardrails or regulatory controls. This is best exemplified by the Executive Order 14179, titled "Removing Barriers to American Leadership in Artificial Intelligence"[22] adopted by President Donald J. Trump in his second administration.[23]

Executive Order 14179 "revokes certain existing AI policies and directives that act as barriers to American AI innovation, clearing a path for the United States to act decisively to retain global leadership in artificial intelligence". A key focus of the executive order is "America's global AI dominance"[24] and the promotion of artificial intelligence as a tool for national security and economic competitiveness. By adopting this executive order, the Trump Administration nullified the guardrails and protections included within the President Joseph R. Biden's Executive Order 14110 "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence".[25] Executive Order 14110 had established eight guiding principles for the development and use of artificial intelligence technologies, including that artificial intelligence be safe and secure, innovation be responsible, and that artificial intelligence protect Americans' privacy and civil liberties.[26]

Under the Trump administration, Executive Order 14179 directed the creation of an "Artificial Intelligence Action Plan". In July 2025, the "Winning the Race, AMERICA'S AI ACTION PLAN" was released.[27] The Plan lays out three pillars: (1) Accelerate AI Innovation; (2) Build American AI Infrastructure; and (3) Lead in International AI Diplomacy and Security.

Regarding regulation, the Plan states "[t]he United States needs to innovate faster and more comprehensively than our competitors in the development and distribution of new AI technology across every field, and dismantle unnecessary regulatory barriers that hinder the private sector in doing so".[28] The Plan continues that "AI is far too important to smother in bureaucracy at this early stage, whether

---

[22] *Removing Barriers To American Leadership In Artificial Intelligence*, Executive Order 14179, Jan. 23, 2025, available at https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/

[23] It is important to note that E.O. 14179 was not President Trump's first executive order related to AI. In his first term, President Trump adopted Executive Order 13859, *Maintaining American Leadership in AI*, available at https://www.federalregister.gov/documents/2019/02/14/2019-02544/maintaining-american-leadership-in-artificial-intelligence. This executive order also focused on the importance of remaining a dominant player in the development of AI technologies for both economic and national security reasons.

[24] *Ibid.*

[25] *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, Executive Order 14110, Oct. 30, 2023, available at https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence.

[26] *Ibid.*

[27] *Winning the Race America's AI action plan*, July 2025, available at https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf (last visited 30/09/2025).

[28] *Ivi,* 1.

at the state or Federal level".[29] Specifically addressing any state-level regulation, the Plan explains: "The Federal government should not allow AI-related Federal funding to be directed toward states with burdensome AI regulations that waste these funds, but should also not interfere with states' rights to pass prudent laws that are not unduly restrictive to innovation".[30]

The Trump administration's innovation-first approach to artificial intelligence is further buttressed by the recently adopted "One, Big Beautiful Bill", introduced in May 2025.[31] During its drafting, this bill included a moratorium on artificial intelligence regulation at the state-level or by any political division. Specifically, the Act stated that 'no State or political subdivision thereof may enforce, during the 10-year period beginning on the date of the enactment of this Act, any law or regulation of that State or a political subdivision thereof limiting, restricting, or otherwise regulating artificial intelligence models, artificial intelligence systems, or automated decision systems entered into interstate commerce'.[32]

Ultimately, this moratorium was removed from the final bill before it was passed.[33] However, it echoes the theme of the AI Action Plan: the Trump administration is focused on artificial intelligence innovation and not creating regulation at either the federal or state level.

The U.S. Federal government's lack of AI regulation on the development and deployment of AI needs to be contrasted with a more proactive regulatory approach to the development of certain artificial hardware and the economic tariffs being set by the Trump administration. These can be seen as regulating artificial intelligence, but in a more protectionist approach versus the creation of trustworthy artificial intelligence as seen in the EU.

Under both the Biden and the Trump Administrations, the U.S. has restricted the sale of semiconductor and chip technologies to certain countries. Using export control measures, the Biden Administration first implemented sweeping export control restrictions in 2022 for the sale of advanced U.S. semiconductors and technologies to China. These controls were further enhanced in 2023 with additional restrictions on the sale of these technologies to China.[34] The Trump administration has continued these restrictions under his second administration, and even explored the requirement that for any chips made for non-U.S. users, there must be a chip made for a U.S. use.[35] These semiconductor technologies are key for the development of large scale artificial intelligence modeling that can compete on the global stage.[36]

---

[29] *Ivi,* 3.

[30] *Ibid.*

[31] *H.R.1 – One Big Beautiful Bill Act*, available at https://www.congress.gov/bill/119th-congress/house-bill/1/text.

[32] Sec. 43201(c) of *H.R.1 – One Big Beautiful Bill Act* reported in House on May 20, 2025.

[33] *State AI Regulation Survived a Federal Ban. What Comes Next?*, in *Carnegie Endowment*, 2025, available at https://carnegieendowment.org/emissary/2025/07/ai-congress-bill-state-ban-what-next.

[34] *Commerce Strengthens Export Controls to Restrict China's Capability to Produce Advanced Semiconductors for Military Applications*, Bureau of Industry & Security, Office of Congressional and Public Affairs, Dec. 2, 2024, available at https://www.bis.gov/press-release/commerce-strengthens-export-controls-restrict-chinas-capability-produce-advanced-semiconductors-military (last visited 30/09/2025).

[35] *U.S. plans 1:1 chip production rule to curb overseas reliance*, *WSJ reports*, in *Reuters*, Sep. 26, 2025, available at https://www.reuters.com/world/us/us-plans-mandate-11-ratio-domestically-manufactured-imported-chips-wsj-reports-2025-09-26/ (last visited 30/09/2025).

[36] *The Intersection of AI and Semiconductors*, Microchip U.S.A, available at https://www.microchipusa.com/industry-news/the-intersection-of-ai-and-semiconductors-advancements-implications-and-future-opportunities (last visited Sep 30, 2025).

Building on this approach, in August 2022, the U.S. passed the CHIPS and Science Act,[37] which authorized $280 billion to boost domestic research and manufacturing of semiconductors in the United States. The Act is intended to create a more controlled supply chain around the key technologies needed to promote American dominance in the artificial intelligence industry.

Contrary to the more hands-off regulatory approach for the deployment of artificial intelligence technologies into society, the U.S. has taken a much more proactive approach to regulating the tools needed to develop AI technologies: creating barriers to the sale of certain technologies outside of the U.S. (and specifically, China) and encouraging, through huge financial incentives, maintaining AI research and development within the U.S.. These examples highlight the complex role of regulation of artificial intelligence at the U.S. federal level.

2.2.2. Existing U.S. Federal Regulations That Impact Certain AI Use Cases

The lack of a specific Federal artificial intelligence regulation on the use of artificial intelligence technologies does not mean that there are no regulations that could impact specific use cases of artificial intelligence within the U.S. Existing laws such as the Health Insurance Portability and Accountability Act (HIPAA)[38] and the Federal Trade Commission Act (FTC Act)[39] are being leveraged in certain instances to create protections around the deployment of artificial intelligence.

The U.S. Department of Health and Human Services (HHS), through its Office for Civil Rights (OCR), is charged with enforcing HIPAA. In this role, OCR issued a "Dear Colleague" letter in January 2025 related to "Ensuring Nondiscrimination Through the Use of Artificial Intelligence and Other Emerging Technologies".[40] In this January 2025 OCR Letter, OCR clearly calls out the need to balance the value that the healthcare industry can receive from the use of artificial intelligence technologies, with the potential risks to patients for discriminatory treatment. OCR adopted final rule implementing Section 1557 in 2024, which "prohibits discrimination on the basis of race, color, national origin, age, sex, and disability in health programs or activities that receive Federal financial assistance from HHS, health programs or activities established under Title I, such as State-based Exchanges, and HHS-administered health programs or activities, including the Federally-facilitated Exchanges".[41]

The January 2025 OCR Letter clarifies that these anti-discrimination requirements apply equally to any "patient care decision support tool" including those tools that leverage artificial intelligence. In short, OCR is taking the position that the HIPAA regulation, and its corresponding protections, are technology agnostic, and will apply equally to artificial intelligence as well as other technologies.

Turning to the Federal Trade Commission (the FTC), the FTC is, in essence, a consumer protection and antitrust regulator.[42] The FTC Act, Section 5, which is the main governing authority for the FTC, empowers the FTC to investigate and prevent unfair methods of competition, and unfair or deceptive

[37] 136 Stat. 1366, available at https://www.congress.gov/bill/117th-congress/house-bill/4346.
[38] Pub. L. No. 104-191, 110 Stat. 1936 (1996).
[39] 15 U.S.C. § 45.
[40] Department of Health & Human Services, Letter re: *Ensuring Nondiscrimination Through the Use of Artificial Intelligence and Other Emerging Technologies*, January 10, 2025 (hereinafter, "January 2025 OCR Letter").
[41] January 2025 OCR Letter, 1-2.
[42] Federal Trade Commission, *What the FTC Does*, available at https://www.ftc.gov/news-events/media-resources/what-ftc-does (last visited 30/09/2025).

acts or practices affecting commerce.[43] The FTC is leveraging both competition and unfair or deceptive acts in the context of artificial intelligence to protect consumers against any harm created by artificial intelligence technologies.

In January 2025, the FTC provided a detailed overview of its position on artificial intelligence and potential consumer harms that may result from the use of artificial intelligence.[44] In this overview, the FTC makes clear that AI is not exempt from existing laws:

> Because there is no AI exemption from the laws on the books, firms deploying these AI systems and tools have an obligation to abide by existing laws, including the competition and consumer protection statutes that the FTC enforces. FTC staff can analyze whether these tools violate people's privacy or are prone to adversarial inputs or attacks that put personal data at risk. We can also scrutinize generative AI tools that are used for fraud, manipulation, or non-consensual imagery, or that endanger children and others. We can consider the impacts of algorithmic products that make decisions in high-risk contexts such as health, housing, employment, or finance. Those are just a few examples, but the canvas is large.[45]

This is a strong statement reminding all developers and deployers of AI that they are subject to regulatory oversight, even if no stand-alone AI law currently exists.

The FTC further lays out a number of factors that businesses should take into account when leveraging AI in their operations, including:

1. Taking necessary steps to prevent harm before and after deploying a product.
2. Taking preventative measures to detect, deter, and halt AI-related impersonation, fraud, child sexual abuse material, and non-consensual intimate imagery.
3. Avoiding deceptive claims about AI tools that result in people losing money or put users at risk of harm.
4. Ensuring privacy and security by default.[46]

In February 2024, former Chair and Commissioner, Lina M. Khan, provided remarks on the intersection of competition law and artificial intelligence technologies.[47] In her remarks, Ms. Khan emphasized the need for the FTC to establish "rules of the road for AI".[48] She laid out four areas where the FTC is focusing on artificial intelligence technologies from a competition lens. First, the FTC is reviewing any existing or emerging bottlenecks across the AI stack. History shows that firms that capture control over key inputs or distribution channels can use their power to exploit those bottlenecks, extort customers, and maintain their monopolies".[49] Second, the FTC is "focused on how business models drive incentives"

[43] *Ibid.*; *Federal Trade Commission Act* (FTC Act), 15 U.S.C. §§ 41-58, at §45(a)(1).
[44] Federal Trade Commission, *AI and the Risk of Consumer Harm*, January 3, 2025.
[45] *Ibid.*
[46] *Ibid.*
[47] Federal Trade Commission, *A few key principles: An excerpt from Chair Khan's Remarks at the January Tech Summit on AI*, February 8 2024, available at https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2024/02/few-key-principles-excerpt-chair-khans-remarks-january-tech-summit-ai (last visited 30/09/2025).
[48] *Ibid.*
[49] *Ibid.*

and that "[t]he drive to refine your algorithm cannot come at the expense of people's privacy or security, and privileged access to customers' data cannot be used to undermine competition".[50]
Third, the FTC is aiming to align "liability with capability and control. This requires looking upstream and across layers of the AI stack to pinpoint which actor is driving or enabling the lawbreaking".[51] Fourth, the FTC, recognizing the uniqueness created by AI, is looking to create "effective remedies that establish bright-line rules on the development, use, and management of AI inputs".[52] For example, in the FTC's opinion "some data is simply off the table for training models".[53]
Taken together, the FTC's guidance to date demonstrates that at the federal level in the U.S., there is not a complete absence of any regulatory oversight regarding artificial intelligence. The challenge is that the FTC's authority is somewhat limited to a consumer protection or competition lens. And, often, there are business use cases of artificial intelligence that may fall outside of the scope of the FTC's authority to regulate, leaving certain areas exposed to the risk of little federal regulation or oversight.

2.2.3. The NIST AI Risk Management Framework as a Non-Regulatory Response

Beyond regulation, in January 2023, the National Institute of Standards and Technology (NIST),[54] part of the U.S. Department of Commerce, released the Artificial Intelligence Risk Management Framework (AI RMF), "a resource to the organizations designing, developing, deploying, or using AI systems to help manage the many risks of AI and promote trustworthy and responsible development and use of AI systems".[55] While the framework is not a regulation, it does represent the most comprehensive attempt at the federal level, to date, to articulate a structured, socio-technical approach to governing artificial intelligence risks.
The AI RMF, as with all NIST guidance, provides guidance that is "intended to be *voluntary*, rights-preserving, non-sector-specific, and use-case agnostic, providing flexibility to organizations of all sizes and in all sectors and throughout society to implement the approaches in the Framework".[56] It creates a risk-based approach, not unlike the EU AI Act, trying to both "minimize anticipated negative impacts of AI systems and identify opportunities to maximize positive impacts".[57] The AI RMF attempts to provide practical guidance for managing a wide spectrum of harms artificial intelligence systems may produce—ranging from bias and discrimination to safety failures and privacy intrusions. The AI RMF is not limited to technical safeguards; it includes organizational governance, accountability, and the embedding of legal and ethical considerations into the lifecycle of AI systems.
Even though the AI RMF is not a regulation, it still could impact the development of artificial intelligence technologies and at least provide a framework to assess the technologies against certain guardrails and controls. As in other areas of emerging technology, voluntary standards often evolve into benchmarks

[50] *Ibid.*
[51] *Ibid.*
[52] *Ibid.*
[53] *Ibid.*
[54] NIST is charged with creating standards in various fields of science and technology.
[55] *Artificial Intelligence Risk Management Framework* (AI RMF 1.0), NIST AI 100-1, January 2023, available at https://doi.org/10.6028/NIST.AI.100-1 (last visited 30/09/2025).
[56] *Ivi*, 2 (emphasis in original).
[57] *Ivi*, 4.

for regulatory expectations, procurement requirements, and even judicial determinations of reasonable care. In the absence of a comprehensive federal artificial intelligence statute, the AI RMF offers a soft law infrastructure that may guide U.S. federal agencies as well as the private sector, helping to provide a foundation for any industry self-regulation or a standard of reasonableness for artificial intelligence.

2.2.4. The U.S. States Attempt to Create Artificial Intelligence Regulatory Initiatives

While artificial intelligence regulation at the federal level in the U.S. remains sparse, the U.S. states continue to push forward with artificial intelligence regulations, either in the form of stand-alone regulations, such as Colorado's AI Act[58] or the Texas Responsible Artificial Intelligence Governance Act,[59] or by leveraging the growing number of comprehensive state-level privacy laws to implement requirements around the use of personal information in automated decision technologies. California is the best example of this second approach. This Article will review each in turn.

Colorado adopted the first comprehensive artificial intelligence regulation in the U.S. in May 2024: An Act Concerning Consumer Protections in Interactions with Artificial Intelligence Systems (the Colorado AI Act). Effective on January 1, 2026, the Colorado AI Act lays out requirements for developers and deployers of artificial intelligence in Colorado.

Drawing parallels to the EU AI Act, the Colorado AI Act defines an "artificial intelligence system" as "any machine-based system that, for any explicit or implicit objective, infers from the inputs the system receives how to generate outputs, including content, decisions, predictions, or recommendations, that can influence physical or virtual environments".[60] However, the Colorado AI Act focuses exclusively on "high-risk artificial intelligence systems", defined as "any artificial intelligence system that, when deployed, makes, or is a substantial factor in making, a consequential decision".[61]

Breaking down that definition further, the Colorado AI Act defines the terms "substantial factor" and "consequential decision". Substantial factor is "a factor that: (i) assists in making a consequential decision; (ii) is capable of altering the outcome of a consequential decision; and (iii) is generated by an artificial intelligence system".[62] This can include "any use of an artificial intelligence system to generate any content, decision, prediction, or recommendation concerning a consumer that is used as a basis to make a consequential decision concerning the consumer".[63] Consequential decision is defined as "a decision that has a material legal or similarly significant effect on the provision or denial to any consumer of, or the cost or terms of: (a) education enrollment or an education opportunity; (b) employment or an employment opportunity; (c) a financial or lending service; (d) an essential government service; (e) health-care services; (f) housing; (g) insurance; or (h) a legal service".[64]

---

[58] An Act concerning consumer protections in interactions with artificial intelligence systems, Colorado Senate Bill 24-205 (SB 24-205) (hereinafter *Colorado AI Act*).

[59] An Act relating to regulation of the use of artificial intelligence systems in this state; providing civil penalties, Texas House Bill 149 (H.B. 149) (hereinafter *Texas AI Act*), available at https://capitol.texas.gov/tlodocs/89R/billtext/pdf/HB00149F.pdf#navpanes=0.

[60] *Colorado AI Act*, 6-1-1701(2).

[61] *Colorado AI Act*, 6-1-1701(9)(a).

[62] *Colorado AI Act*, 6-1-1701(11)(a).

[63] *Colorado AI Act*, 6-1-1701(11)(b).

[64] *Colorado AI Act*, 6-1-1701(3).

In many ways, the Colorado AI Act is very limited in scope both in terms of needing a "high-risk" use of artificial intelligence combined with its use in eight delineated areas of society. Lower risk uses of artificial intelligence are not governed by the law, and as such, remain unregulated unless they fall into another regulatory regime (such as healthcare or consumer data privacy laws).

In June 2024, Texas adopted the Texas Responsible Artificial Intelligence Governance Act, which is primarily aimed at "facilitat[ing] and advanc[ing] the responsible development and use of artificial intelligence systems".[65] The Texas AI Act is intended to (1) protect individuals and groups of individuals from known and reasonably foreseeable risks associated with artificial intelligence systems; (2) provide transparency regarding risks in the development, deployment, and use of artificial intelligence systems; and (3) provide reasonable notice regarding the use or contemplated use of artificial intelligence systems by state agencies.[66]

The Texas AI Act applies broadly to any person who: "(1) promotes, advertises, or conducts business in this state; (2) produces a product or service used by residents of this state; or (3) develops or deploys an artificial intelligence system in this state".[67] It creates a "regulatory sandbox" that "enables a person to obtain legal protection and limited access to the market in this state to test innovative artificial intelligence systems without obtaining a license, registration, or other regulatory authorization".[68] In this way, the Texas AI Act is attempting to strike a balance between protecting individuals from higher risk uses of AI while also allowing for innovation in this more fluctuating time of artificial intelligence development.

California has taken a different approach to regulating artificial intelligence than Colorado and Texas. First, instead of focusing on a more comprehensive artificial intelligence regulation, it has attempted (with marginal success) in adopting smaller, more specific artificial intelligence regulation in nuanced areas. For example, in January 2025, eighteen (18) laws related to AI went into effect in California. These laws fall, generally, into six categories.

First, California enacted two (2) general laws that create a standard definition of artificial intelligence in California and outline documentation requirements for the training and development of artificial intelligence models.[69] Second, California enacted eight (8) laws that are intended to protect individuals against certain use-cases of artificial intelligence, including protecting performers' rights, extending existing laws to protect against child sexual abuse materials with AI-generated materials and the use of deceptive AI-generated content in the political context, and prohibiting the use of non-consensual deepfake pornography.[70]

---

[65] *Texas AI Act*, Sec. A551.003(1).
[66] *Texas AI Act*, Sec. A551.003(1).
[67] *Texas AI Act*, Sec. A551.002.
[68] *Texas AI Act*, Sec. A553.051.
[69] AB 2885, available at https://legiscan.com/CA/text/AB2885/id/3020074 and AB 2013, available at https://legiscan.com/CA/text/AB2013/id/3019237 (last visited 30/09/2025).
[70] AB 1831 (child pornography laws), AB 1836 (unauthorized use of digital replicas of deceased persons without consent), AB 2602 (unauthorized use of digital replicas), AB 2655 (materially deceptive election-related content), AB 2839 (deceptive AI generated content in election advertisements), AB 2355 (clear disclosures on political advertisements for AI generated content), SB 926 (criminalizing non-consensual deep fake pornography) and SB 981 (reporting tools for social media companies).

Third, California adopted two laws that directly address the use of artificial intelligence in the healthcare context. AB 3030 requires that healthcare providers disclose to patients the use of artificial intelligence technologies in patient communications and provide patients with instructions on how to reach out to a human healthcare provider.[71] SB 1120 requires that only physicians, and not artificial intelligence technologies, can make final decisions in relation to a patient's care and treatment.[72]

Fourth, California adopted two laws that relate to their prior existing privacy legislation, the California Consumer Privacy Act (CCPA),[73] clarifying the intersection of data privacy and artificial intelligence technologies. SB 1223 clarifies that neural data is considered sensitive personal information under the CCPA, and as such, receives heightened protections under the privacy law.[74] This is relevant in the artificial intelligence context as it will require express consent from a user to use neural data with any artificial intelligence technologies. AB 1008 expressly states that AI-generated content could be considered personal information if it can be used to identify an individual.[75] As such, the consumer data rights outlined under the CPPA apply equally to artificial intelligence systems that may be ingesting personal information of California consumers and requires that deployers of artificial intelligence build in the capability to respond to those data rights as required under the CCPA.

Fifth, California adopted two laws that relate to AI transparency. SB 942 requires that "covered providers" of artificial intelligence technologies provide users with free artificial intelligence detection tools to allow users to identify whether content was generated or altered by AI technologies.[76] This transparency initiative is intended to give users control to identify if what they are viewing online or any context is real or created by artificial intelligence. AB 2905 relates to the use of artificial voices in auto-dialing technologies.[77] Specifically, companies must notify an individual, in a real voice, that the following message was created by an AI-generated voice.

Finally, the last group of enacted laws relate to the use of artificial intelligence technologies in the government and school settings. SB 896 requires that the California government assess the impact of AI on its critical infrastructure and to provide annual reports to the California legislature.[78] AB 2876 requires that AI-literacy be incorporated into the K-12 education curriculum.[79] And, SB 1288 requires the establishment of a working group to assess the safe and effective use of AI in public schools.[80]

---

[71] AB 3030, available at https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240AB3030 (last visited 30/09/2025).

[72] SB 1120, available at https://legiscan.com/CA/text/SB1120/id/3023335 (last visited 30/09/2025).

[73] Cal. Civ. Code § 1798.100 et seq., as amended.

[74] SB 1223, available at https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB1223 (last visited 30/09/2025).

[75] AB 1008, available at https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240AB1008 (last visited 30/09/2025).

[76] SB 942, available at https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB942 (last visited 30/09/2025).

[77] AB 2905, available at https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240AB2905 (last visited 30/09/2025).

[78] SB 896, available at https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB896 (last visited 30/09/2025).

[79] AB 2876, available at https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240AB2876 (last visited 30/09/2025).

[80] SB 1288, available at https://legiscan.com/CA/text/SB1288/id/2981356 (last visited 30/09/2025).

Taken together, these newly enacted laws demonstrate a concerted effort to address artificial intelligence in California, even if it creates a more patchwork regulatory approach. In addition to these newly enacted laws, the California Privacy Protection Agency (CPPA) recently adopted regulations related to automated-decision making technology (ADMT).[81] The CPPA authority is limited to enforcing and providing guidance under the CPPA, so its regulations inherently relate to the overlap between ADMT and personal information.

The ADMT regulations define ADMT as "any technology that processes personal information and uses computation to replace human decision making or substantially replace human decision making".[82] By its very definition, it is limited to processing of personal information, an area already regulated within California. The ADMT Regulations create rights for consumers to (1) request to access information related to the use of ADMT technologies and the impact on the consumer; and (2) appeal any decision of the business to use ADMT for a significant decision.[83]

The ADMT Regulations outline certain requirements for any business that is using ADMT. First, the business must provide a "Pre-use notice" of any ADMT technologies to the consumer that also includes a link for the consumer to opt-out of the use of ADMT for processing the consumer's personal information.[84] These "Pre-use notices" must be "presented prominently and conspicuously to the consumer at or before the point when the business collects the consumer's personal information that the business plans to process using ADMT".[85]

Further, businesses are required to conduct risk assessments for any use of ADMT for a "significant decision concerning a consumer" or when the business will use a consumer's personal information for training an ADMT to make a significant decision regarding consumers.[86] For these risk assessments regarding ADMT, the business must identify "(i) [t]he logic of the ADMT, including any assumptions or limitations of the logic; and (ii) [t]he output of the ADMT, and how the business will use the output to make a significant decision".[87] And, the business must document the policies, procedures, and training in place "to ensure that the business's ADMT works as intended for the business's purpose and does not unlawfully discriminate based upon protected characteristics".[88] These risk assessments apply to businesses that are leveraging ADMT as well as suppliers of ADMT, who must provide sufficient

[81] California Privacy Protection Agency, *Proposed Regulations on CCPA Updates, Cybersecurity Audits, Risk Assessments, Automated Decisionmaking Technology (ADMT), and Insurance Companies*, available at https://cppa.ca.gov/regulations/ccpa_updates.html (hereinafter *ADMT Regulations*) (last visited 30/09/2025). These ADMT Regulations were adopted by the CPPA board during its July 24, 2025 meeting and approved by the California Office of Administrative Law (OAL) on September 23, 2025. *See*, CPPA, *California Finalizes Regulations to Strengthen Consumers' Privacy*, September 23, 2025, available at https://cppa.ca.gov/announcements/2025/20250923.html (last visited 30/09/2025).

[82] CCPA, *Modified text of proposed regulations*, available at https://cppa.ca.gov/regulations/pdf/ccpa_updates_cyber_risk_admt_mod_txt_pro_reg.pdf (last visited 30/09/2025).

[83] *ADMT Regulations*

[84] *ADMT Regulations*, 20.

[85] *ADMT Regulations*, 102.

[86] *ADMT Regulations,* 84-85.

[87] *ADMT Regulations*, 89.

[88] *ADMT Regulations*, 92.

information to allow a business using the ADMT to conduct a sufficient risk assessment as required under the Regulations.[89]

Overall, while the U.S. federal government is not taking proactive regulatory measures, the U.S. states are filling that void with a variety of approaches to artificial intelligence regulation. While the state-focused approach provides a unique opportunity to 'test' different regulatory approaches, this will prove challenging with artificial intelligence, which is impactful across borders and the globe, and so heavily reliant on large multinational companies for its development and deployment.

3. A Comparison of the Regulatory Approaches to Artificial Intelligence in the EU and the U.S.

Artificial intelligence regulation stands at a unique tipping point, bringing in legal components, geopolitical risks, and economic factors. The comparison between the approach between the EU and the U.S. is illustrative of this decisive moment. On the one hand, the U.S. is home to three out of the four most dominant AI models currently available: OpenAI with its ChatGPT; Google's Gemini; and Microsoft's Co-Pilot.[90] DeepSeek, developed and deployed out of China, is one of the only main competitors to these U.S. power houses.[91] Additionally, the U.S. is home to Nvidia, one of the dominant developers and manufacturers of semi-conductor chip technology that is a critical component of any of these large-scale AI models.[92]

The EU has none of these factors in play. While there are rapid attempts to encourage the development of artificial intelligence within Europe, it is not yet at a level where it is competing on a global stage.[93] France is an outlier in actively developing global artificial intelligence models.[94] Instead, the EU, in many ways, appears to heading towards a future where it is a buyer and user of artificial intelligence technologies that are developed elsewhere, putting it in a unique position of not having direct sovereignty over these technologies, but being highly impacted by their development and guardrails (or lack thereof).

These geopolitical tensions cannot be overstated. In the United States, the current regulatory stance toward the technology sector reflects an ongoing tension between preserving space for innovation and responding to growing demands for oversight, particularly in areas like artificial intelligence, data privacy, and technology governance. Federal policy, particularly under the Biden administration, has leaned toward a 'precision regulation' approach-seeking to target high-risk technologies and use cases without stifling broader innovation. This is evident in initiatives like the 2023 AI Executive Order, which emphasized safety, transparency, and equity, while delegating much of the implementation to sector-specific agencies and voluntary frameworks.

---

[89] *Ibid.*

[90] N. Maslej, L. Fattorini, R. Perrault, Y. Gil, V. Parli, N. Kariuki, E. Capstick, A. Reuel, E. Brynjolfsson, J. Etchemendy, K. Ligett, T. Lyons, J. Manyika, J.C. Niebles, Y. Shoham, R. Wald, T. Walsh, A. Hamrah, L. Santarlasci, J.B. Lotufo, A. Rome, A. Shi, S. Oak, *The AI Index 2025 Annual Report*, Section 1.3, AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, April 2025, *available at* https://hai.stanford.edu/assets/files/hai_ai_index_report_2025.pdf (hereinafter, *Standford Report*).

[91] *Ibid.*

[92] Nvidia, Contact Us, available at https://www.nvidia.com/en-us/contact/ (last visited 30/09/2025).

[93] *Stanford Report*, Section 1.3.

[94] *Stanford Report*, Section 1.3.

In many ways, the U.S. states appear to be adopting a similar "precision regulation" approach. The two comprehensive artificial intelligence laws currently enacted, Colorado and Texas, are arguably not even comprehensive in that they are focused on higher risk use cases or very narrow areas of artificial intelligence usage. And, for states where no comprehensive artificial intelligence law is enacted, or even being discussed, they are leveraging privacy laws or general consumer protection laws to create some level of guardrails for AI development and deployment. This is best exemplified by California.

At the same time, there's palpable hesitation in Congress to enact sweeping regulatory regimes that might inadvertently hamper U.S. competitiveness, especially relative to China and the EU. As a result, much of the regulatory momentum has shifted to the states and to agency rulemaking, creating a patchwork environment where companies navigate a complex web of evolving standards and requirements. This federal ambivalence reflects a broader ideological divide-balancing the historical American preference for market-driven growth with the rising awareness that unchecked technological advancement can pose systemic risks to civil liberties, economic equity, and democratic institutions.

4. How to Think About How to Regulate Artificial Intelligence

The question remains: should governments regulate artificial intelligence, even as its impacts are still unknown? Or should artificial intelligence be allowed to innovate with very few restrictions, in a wait-and-see regulatory approach?

To answer this question would require a crystal ball. Instead, it is beneficial, taking into context the EU and U.S. divergent approaches, to instead consider lessons learned from these first influential years, and how to frame out the discussion of creating controls and expectations in the next phase of artificial intelligence, and ways to inform future regulatory decision making.

4.1. Lesson Number One: Existing Laws Do Provide Some Regulatory Protections, Without Even Mentioning Artificial Intelligence

There are examples, both in the U.S. and Europe, of existing regulatory regimes that provide certain protections and requirements in the context of artificial intelligence, beyond any AI-specific regulations. The best example is data privacy. Where personal information is ingested, or anyway used by an artificial intelligence model, in both the U.S. and Europe, there are privacy requirements that will attach to that use.

In Europe, the European Data Protection Board (EDPB), which oversees EU-wide enforcement of the GDPR, issued a statement on the role of data protection authorities in the EU AI Act.[95] In that statement, the EDPB makes clear that is already actively enforcing GDPR in the context of AI technologies:

> In fact, the processing of personal data (which is often strictly intertwined with non-personal data) along the lifecycle of AI systems – and particularly along the lifecycle of those AI systems presenting a high risk to fundamental rights – clearly is (and will continue to be) a core element of the various technologies covered under the umbrella of the AI definition, as enshrined in Article 3(1) AI Act. For

[95] Statement 3/2024 on data protection authorities' role in the Artificial Intelligence Act framework, adopted July 16, 2024, available at https://www.edpb.europa.eu/system/files/2024-07/edpb_statement_202403_dpasroleaiact_en.pdf (hereinafter, *EDPB Statement*).

> these reasons, national data protection authorities (hereinafter DPAs) have been active with regard to these technological developments9 and the EDPB, which has closely followed the legislative process regarding the AI Act10, has already initiated the examination of its (multifaceted) interplay with EU data protection law.[96]

The statement further clarifies that "whenever a general-purpose AI model or system entails the processing of personal data, it may fall – like any other AI system – under the supervisory remit, as applicable, of the relevant national DPAs (also cooperating according to Chapter VII of the GDPR) and of the EDPS (when it falls under the EUDPR)".[97] As such, even without the EU AI Acts adoption, artificial intelligence technologies are already subject to at least the GDPR when operating within or available to Europeans.

Similarly, in the U.S., as outlined in Section II, both the OCR, under HIPAA, and the FTC are actively leveraging their respective regulatory authority to oversee the use of artificial intelligence in their areas of authority. OCR issued guidance for instances where artificial technologies are used in healthcare context or leveraging protected health information in any capacity. The FTC made clear, and continues to make clear, that any commercial use of artificial intelligence must not result in unfair or deceptive trade practices. Additionally, U.S. states are using their own regulations, such as comprehensive data privacy laws and consumer protection laws, to require certain protections in the use of artificial intelligence. Combined, these measures negate a claim that artificial intelligence is completely unregulated in the U.S. In fact, it is subject to certain regulatory oversight, just not with a stand stand-alone artificial intelligence regulation.

In both the EU and the U.S., regulators should maximize the ability to leverage the existing legal infrastructures to place guardrails on the development of artificial intelligence technologies. While these may leave gaps in certain industries or use cases, they are already in place, today, making them the most effective path forward to oversee certain aspects of artificial intelligence.

4.2. Lesson Number Two: Technical Standards and Controls May Offer a Middle Ground, With Flexibility and Ease of Adoption

Technical standards and controls are a policy option that avoids the burdens and drawn-out processes of regulation but create expectations around artificial intelligence. On the U.S. side, NIST best exemplifies this approach with the AI RMF. Passing legislation in Congress would require political and strategic hurdles, no more so than in the current political environment. To avoid those challenges, voluntary technical controls and standards can fill the gap and create a de facto baseline of reasonableness.

In addition to NIST, the International Organization for Standardization (ISO) developed numerous standards addressing artificial intelligence.[98] The ISO standards address a variety of aspects of artificial intelligence, including key concepts and terminology,[99] management of artificial intelligence,[100] AI

[96] *Ivi,* 4.
[97] *Ivi,* 14.
[98] ISO, *Artificial Intelligence*, available at https://www.iso.org/sectors/it-technologies/ai (last visited 30/09/2025).
[99] ISO/IEC 22989, *Information technology — Artificial intelligence — Artificial intelligence concepts and terminology,* 2022.

system impact assessments,[101] artificial intelligence risk management,[102] and AI system life-cycle processes.[103] ISO standards, unlike NIST, are auditable and provide certifications of compliance, requiring that organizations demonstrate compliance to third-parties in order to affirmatively state compliance with the ISO standard.

While neither the NIST nor ISO standards are mandatory from a regulatory perspective, they are often used in the private sector for businesses to demonstrate a certain level of maturity in reasonable controls in the use of technologies. For example, businesses will often expressly state compliance with one (or both) of these standards in contracts, making voluntary standards from a regulatory standpoint now mandatory from a contractual standpoint. As more businesses are beginning to use and integrate artificial intelligence into their daily operations, business are increasingly requiring artificial intelligence developers and vendors to enter into contracts that use one of these standards to set the baseline expectations for how the tools will function and what protections will be in place for a business. As such, where regulation may lag behind, the private sector market forces may push artificial intelligence technologies to adopt certain controls and safeguards due to customer demands and contractual requirements.

4.3. Lesson Number Three: The Development of Artificial Intelligence Is Isolated to Only a Few Regions in the World

As artificial intelligence continues to develop, the U.S. and China are dominating the development of these tools and technologies. With the clustering of artificial development, and the technological infrastructure to support that development, into such few areas, it becomes even more important for these regions to create a harmonized approach to overseeing the development and use of artificial intelligence tools.

The geographic concentration of artificial intelligence development in the U.S. and China[104] will profoundly shape the regulatory landscape in both the U.S. and Europe. In the U.S., the dominance of domestic artificial intelligence companies creates strong economic and geopolitical incentives to maintain a more permissive regulatory environment. Lawmakers, especially at the federal level, appear particularly focused on the fear that stringent rules could weaken the global competitiveness of U.S. tech companies vis-à-vis their Chinese counterparts. As a result, U.S. regulation is likely to remain sectoral, fragmented, and innovation-friendly, with an emphasis on voluntary standards, soft law, and targeted interventions rather than sweeping statutory frameworks. The concentration of corporate power in a handful of American firms also ensures that industry voices exert significant influence over the regulatory agenda, reinforcing this cautious approach.

In Europe, by contrast, the absence of globally competitive artificial intelligence solutions fuels a different set of incentives. European regulators, recognizing their comparative weakness in artificial intelligence innovation, have positioned themselves as global leaders in artificial intelligence

[100] ISO/IEC 42001, *Information technology — Artificial intelligence — Management system*, 2023.
[101] ISO/IEC 42005, *Information technology — Artificial intelligence (AI) — AI system impact assessment*, 2025.
[102] ISO/IEC 23894, *Information technology — Artificial intelligence — Guidance on risk management*, 2023.
[103] ISO/IEC 5338, *Information technology — Artificial intelligence — AI system life cycle processes*, 2023.
[104] *See*, *supra*, note 90.

governance, much as Europe did with data protection. By setting ambitious standards — such as the EU AI Act — the EU seeks to exercise 'normative power', exporting its regulatory model extraterritorially and shaping the practices of U.S. and Chinese firms that wish to access the European market. The reliance on external artificial intelligence technologies also magnifies European concerns about sovereignty, dependence, and the protection of fundamental rights, all of which drive a more precautionary, rights-oriented regulatory framework.

These unique dynamics mean that the global approach to artificial intelligence regulation will be shaped by a triangular interplay: U.S. regulators protecting domestic innovators, EU regulators projecting values-based governance outward, and China pursuing state-led strategies that blend technological development with political control.[105] The result is not a harmonized system, but a fragmented and competitive regulatory environment, in which Europe attempts to wield rule-making authority, the U.S. prioritizes innovation and competitiveness, and both must ultimately reckon with the reality that the technological frontier is being set largely outside of Europe. This asymmetry will continue to shape both the form and ambition of regulatory initiatives across the Atlantic.

5. Conclusion

There is no right answer to when and how to regulate any emerging technology – and artificial intelligence is no different. However, artificial intelligence does pose unique challenges by its very nature as an incredible powerful tool and its rapid adoption by users across the globe. The regulatory trajectories of the EU and the U.S. reflect not merely divergent legal traditions but fundamentally different philosophies of technology governance. The EU, through its comprehensive EU AI Act, embraces a precautionary and risk-based framework that prioritizes fundamental rights and harmonized obligations across member states. By contrast, the U.S. leans toward sector-specific and innovation-driven approaches, emphasizing guidance, self-regulation, and adaptive enforcement rather than an overarching statutory regime. Each system responds to its own institutional logic: the EU's deep-rooted emphasis on rights protection and market integration, and the U.S.'s preference for flexibility, competitiveness, and decentralized oversight.

Yet, these differences also underscore valuable lessons. From the EU, policymakers can see the benefits of legal certainty, uniformity, and proactive safeguards in cultivating public trust. From the U.S., the value of flexibility, experimentation, and avoiding premature over-regulation becomes evident, ensuring that technological innovation is not unduly stifled. The juxtaposition highlights the tension between fostering innovation and protecting society, a balance that all jurisdictions must carefully navigate.

Ultimately, the future of artificial intelligence regulation will likely require hybridization: combining the EU's structured rights-based protections with the U.S.'s adaptive and sectoral responsiveness. As artificial intelligence becomes a global technology, transatlantic convergence — or at least interoperability — will be crucial to avoid fragmentation, ensure accountability, and preserve both innovation and human dignity. The comparative experience suggests that the most effective regulatory path forward lies not in rigid adherence to one model, but in the dialogue between them.

[105] ZHANG, A. HUYUE, *The Promise and Perils of China's Regulation of Artificial Intelligence*, in *Columbia J. Trans.*, 2025, 5-6.

Essays

AI and Personal Data Regulation: From Public Authority Enforcement to Civil Liability Law

Erwann Picart-Cartron*

ABSTRACT: The administrative procedure before the supervisory authority and the possibility to act under liability law are remedies both available under AI act and GDPR. On one hand, the supervisory authority has an ambivalent role. It operates at the intersection of market regulation and the protection of fundamental rights, combining both ex-ante and ex-post powers. On the other hand, liability law must not be undervalued in ensuring the enforcement of these regulations, especially given its preventive (prophylactic) function. Analyzing these remedies allows to highlight how the GDPR will be a key component in ensuring the effectiveness of the AI Act. Conversely, it illustrates how artificial intelligence provides a new perspective on certain provisions of data protection law.

KEYWORDS: personal data protection; artificial intelligence; regulation; civil liability; supervisory authorities

SUMMARY: 1. Introduction: the guiding principles of the AI Act and the GDPR – 2. The shortcomings in the role of supervisory authorities – 2.1. The mesh of supervisory authorities – 2.2. The uncertainties surrounding sanctions by personal data authorities – 3. The possible remedies available under civil liability – 3.1. The limited scope of civil liability under the AI Act – 3.2. The extent of remedies under the GDPR – 4. Conclusion.

1. Introduction: the guiding principles of the AI Act and the GDPR

Artificial intelligence is an emerging technology, to the point that its true impact and the progress it generates remain difficult to fully assess. Despite these uncertainties, AI is rapidly integrating into every aspect of daily life, valued by individuals for its inference and automation capabilities. While these features can be used to simplify research or create entertaining content, they are also deployed by companies to monitor virtually every dimension of a person's existence, whether in the workplace, on the street, or in private spaces through smartphone activity.

AI thus becomes a powerful tool for collecting and generating personal data. A striking example is Google Vision, which can infer a wide range of personal information from simple images, from objective data such as eye color to sensitive details like religious or political beliefs.[1] This processing often occurs without individuals' awareness, and the data collected are subsequently used to refine user profiling.

* *PhD in law, Contractual Lecturer-Researcher, Associate researcher at IODE Laboratory (Law faculty of Rennes). Mail: erwann.picart@univ-rennes.fr.* *This article was subject to a blind peer review process.*

[1] This website is based on GoogleVision API in order to show the user what can be inferred from a single photo by the software Google Vision: https://theyseeyourphotos.com/.

This dynamic is central to understanding the risks posed by AI: as a novel technology, it amplifies well-known threats to individuals precisely because it autonomously processes data that, when related to an individual, undoubtedly qualify as personal data under General data protection regulation.[2]

For this reason, AI represents a new lens through which to examine the GDPR, offering fresh perspectives on the regulation of personal data.

The widespread adoption of AI and the extensive use of these technologies have prompted the European Commission to adopt the AI Act.[3] However, economic growth, one of the key benefits promised by AI, remains a critical factor in shaping this regulation.[4] The AI Act strikes a balance by treating AI as a product while aiming to regulate its development without stifling innovation or discouraging investment. To achieve this, the AI Act adopts a human-centric approach, ensuring that AI systems remain under human oversight at all times. For individuals, this means they must always be informed when interacting with an AI system and know whom to contact if needed. In this regard, and in many others, the AI Act shares similarities with the GDPR.

Indeed, the GDPR and the AI Act follow a similar logic. In many ways, the AI Act can be seen as the 'new GDPR' as both aim to regulate the development of emerging technologies. This similarity is evident in the extensive number of amendments proposed during discussions in the European Parliament for each text, as well as in the desire to leverage the Brussels Effect.[5] This effect refers to the extraterritorial influence of European legislation, reflecting the European legislator's ambition to set a global benchmark for high standards of protection, whether in personal data (as with the GDPR) or in artificial intelligence (as with the AI Act).

To achieve this, both European regulations rely on a risk-based approach. This approach pursues two objectives. The first is "to prevent risky activities, meaning activities that combine a probability of harm with varying degrees of severity".[6] The provisions of the GDPR thus aim to regulate the flow of personal data so that their processing does not pose disproportionate risks to the data subjects. As for the AI Act, it is structurally based on this approach,[7] as it adopts a classification of AI systems into four categories: prohibited AI practices,[8] high-risk AI systems,[9] and others.[10] This risk-based approach in European regulations introduces a new paradigm that is not purely based on banning practices but rather on

[2] Art. 4, 1), of regulation (EU) 2016/679 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, and repealing Directive 95/46/EC (General Data Protection Regulation, or GDPR).

[3] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) N°167/2013, (EU) N°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, or AI Act).

[4] Recitals 3 and 4 of the AI Act specifically mentioned economic growth linked to development of human centric AIS.

[5] B. ANU, *The Brussels Effect*, in *Northwestern University Law Review*, 107(1), 2012; A. JEAUNEAU, *The Brussels Effect. How the European Union Rules the World*, in *Rev. crit. DIP*, 2021, 525.

[6] A. LATIL, *Digital Law: A Risk-Based Approach*, 2023, 7.

[7] A. LATIL, *Le Règlement relatif à l'intelligence artificielle et l'approche par les risques: application d'une méthode législative structurante*, in *RTD eur.*, 2024, 563.

[8] Art. 5, AI Act.

[9] Art. 6, AI Act.

[10] Art. 51, AI Act.

holding stakeholders accountable.

The accountability principle requires that both the data controller and the AI system operator document all steps taken to comply with applicable regulations. These compliance procedures shape stakeholder practices ex ante by establishing clear expectations. They also enable ex post enforcement, allowing sanctions to be imposed in the event of a proven breach. These two dimensions of accountability, preventive and corrective, are interdependent. If sanctions are insufficient, the principle loses its effectiveness, undermining the regulatory paradigm that supports these frameworks. This is particularly significant because the GDPR and the AI Act may apply concurrently, creating a dual effect: increasing obligations for controllers using AI systems while simultaneously strengthening individuals' rights and means of defense.

The GDPR provides two distinct types of remedies. The first is an administrative remedy, which involves the supervisory authority.[11] This authority plays a unique role in upholding procedural rights. It combines both ex-ante and ex-post capabilities:[12] to oversee personal data processing, the supervisory authority can issue soft law (such as guidelines or recommendations) to assist controllers in achieving compliance.[13] Its ex-post powers come into effect in cases of GDPR violations, encompassing both investigative authority and the power to impose sanctions when necessary.[14]

The second category of remedies is based on the principle of civil liability. Under Article 82 of the GDPR, any data subject who has suffered material or non-material damage due to a breach of the Regulation has the right to claim compensation. This right is conditional on proving both a fault attributable to the controller or processor and the existence of compensable damage.

The AI Act also establishes a similar dichotomy of remedies. While the administrative remedy is directly stipulated within the AI Act itself,[15] civil liability is addressed through an amendment to the European Product Liability Directive.[16] This means that, unlike the GDPR, the liability of an AI controller under the AI Act cannot be directly invoked unless the AI system is deemed defective.

Many questions arise from the similarities between the GDPR and the AI Act, particularly regarding their concurrent application. Since the GDPR's system of remedies has been in force since 2018, it provides a solid basis for analysis. The insights gained from this analysis are invaluable for understanding and anticipating the enforcement of the AI Act. To this end, the paper is based on the French implementation of both regulations.

First, it is possible to identify certain shortcomings in the role of supervisory authorities, particularly in their use of ex-post sanctioning powers (1). From the perspective of data subjects, the sanctioning of GDPR violations serves as a marker of trust in the legal system established by the regulation. However, secondly, if trust in administrative authorities is undermined, individuals may seek alternative remedies, such as civil liability claims against controllers or AI operators, to protect their interests (2).

---

[11] Art. 77, GDPR.

[12] On this matter: M. HERVIEU, *Independent administrative authorities and the renewal of general contract law*, 118, 2012.

[13] Art. 57, GDPR.

[14] Art. 58, GDPR.

[15] Art. 85, AI Act.

[16] Directive (EU) 2024/2853 of the European Parliament and of the Council of 23 October 2024 on liability for defective products and repealing Council Directive 85/374/EEC.

2. The shortcomings in the role of supervisory authorities

In France, the concept of a supervisory authority first emerged with the adoption of the 1978 Data Protection Act (Loi Informatique et Libertés).[17] This law was introduced in a specific digital context, where only public administrations had access to computer systems. At the time, the supervisory authority, originally the Commission Nationale de l'Informatique et des Libertés (CNIL), was established to monitor the use of these systems by the administration, particularly to prevent the profiling of citizens in their interactions with the state.[18]

This foundational role did not disappear with the evolution of French law. However, the adoption of the GDPR in 2018 significantly modified it. The most notable change lies in the introduction of the accountability principle, which requires data controllers to actively ensure that personal data is processed in compliance with GDPR provisions. Under this principle, controllers must document their compliance procedures to prepare for potential audits by the supervisory authority.[19] While this approach is central to the GDPR's regulatory logic, it has also revealed one of the main weaknesses in France's personal data protection framework: the challenge of effectively enforcing accountability in practice.

The AI Act follows a similar logic, extending the supervisory authority's mandate to ensure compliance by AI operators. This continuity underscores the expanding role of supervisory authorities in regulating emerging technologies, while also raising questions about their capacity to address the complexities of AI governance.

However, the cornerstone of this regulatory paradigm remains the sanctioning of controllers or operators in the event of non-compliance with these provisions. In this regard, the concurrent application of the GDPR and the AI Act creates an overlapping web of administrative authorities (2.1), which could complicate the enforcement of the latter regulation. This challenge is particularly acute given that current GDPR enforcement remains unsatisfactory (2.2).

2.1. The mesh of supervisory authorities

The AI act cite multiples supervisory authorities. Some are created in the giron of the European Commission like the AI office,[20] or the European Artificial Intelligence Board.[21] At national level, the AI regulation lay down on existing administrative authorities. They are the notifying authority in charge of "setting up and carrying out the necessary procedures for the assessment, designation and notification of conformity assessment bodies and for their monitoring",[22] or the "market surveillance authority"

[17] M. HERBERT, *Independent administrative authorities: protecting freedoms or social regulation?*, in C.A. COLLIARD, G. TIMSITEDS (sous la dir.), *Independent administrative authorities*, 1988; Conseil d'Etat, *Independent administrative authorities*, public report n. 52, 2001, 257.

[18] J. FRAYSSINET, *The french data protection act 6th January of 1978: a pedagogical and concise overview*, in *R.R.J.*, 2(28) 1987, 191; C. CASTETS-RENARD, *Internet Law: French and European Law*, 2e éd., 2012.

[19] T. DOUVILLE, *Data protection law,* 2023, 265; European Union Agency for Fundamental Rights, *Handbook on European data protection law*, 2018, 194.

[20] Art. 64, AI Act.

[21] Art. 65, AI Act.

[22] Art. 3, 19, AI Act.

nominated under the 2019/1020 regulation concerning market surveillance and compliance of products.[23]

The involvement of market surveillance authorities reveals the dual nature of the EU's approach to AI regulation. While the AI Act aims to promote the development of human-centric AI,[24] its primary focus remains economic: ensuring the security of natural persons while preventing market distortions.[25] As stated in the AI Act, "non-compliant and unsafe products put citizens at risk" and "might distort competition with economic operators selling compliant products within the Union".[26]

Under this framework, market surveillance authorities are tasked with overseeing AI systems that pose risks, treating them like any other product.[27] Their mandate includes monitoring risks "to the health, safety, or fundamental rights of persons".[28] These authorities will assess the compliance of AI systems in the market, assisted by "authorities protecting fundamental rights"[29] already established at the national level.

While data protection authorities are likely to play this role, the European Data Protection Board (EDPB) has also considered their function within the broader scope of market surveillance. This overlap raises questions about the coordination and division of responsibilities between these bodies.[30]

This fragmented regulatory framework creates a mesh of administrative authorities, which risks complicating the oversight of AI's rapid and incessant development. In most EU Member States, including France, multiple administrative bodies are tasked with supervising specific markets.[31] The French government has proposed a "governance scheme for market surveillance authorities", which is currently under parliamentary review.[32] Under this scheme, the Directorate-General for Competition Policy and Consumer Affairs (DGCCRF) serves as the central coordinator. However, responsibilities are divided among several authorities, depending on the category of AI systems identified by the AI Act. For instance, regarding prohibited AI practices under Article 5, the French Data Protection Authority (CNIL) holds jurisdiction over: the use of AI systems for risk assessments of natural persons,[33] the use of AI

[23] Regulation (EU) 2019/1020 of the European Parliament and of the Council of 20 June 2019 on market surveillance and compliance of products and amending Directive 2004/42/EC and Regulations (EC) No 765/2008 and (EU) No 305/2011 (Regulation (EU) 2019/1020 on market surveillance and compliance of products).

[24] Rec. 1, AI Act.

[25] *In this sense: J. Charpenet, Fundamental Rights Put to the Test by the Standardization of Artificial Intelligence*, in *Dalloz IP/IT*, 2025, 591.

[26] Cons. 2, Regulation (EU) 2019/1020 on market surveillance and compliance of products.

[27] *S. Tabani, The European Regulation on Artificial Intelligence: Selected Issues After the First Six Months of the Initial Obligations Coming into Force*, in *Rev. UE*, 2025, 626.

[28] Art. 79, 2, of AI Act refers directly to the art. 3 point 19 of the 2019/1020 regulation on market surveillance and compliance of products.

[29] Art. 77, AI Act.

[30] EDPB, Statement 3/2024 on data protection authorities' role in the Artificial Intelligence Act framework, 16th of july 2024.

[31] Art. 74, 7, AI Act: "By way of derogation from paragraph 6, in appropriate circumstances, and provided that coordination is ensured, another relevant authority may be identified by the Member State as market surveillance authority for the purposes of this Regulation".

[32] The competent authorities for the implementation of the European Regulation on Artificial Intelligence, Directorate General for Enterprises.

[33] AI act, art. 5, §1, d).

systems that create or expand facial recognition databases,[34] the use of AI systems to infer emotions in workplace or educational settings,[35] the use of biometric categorization systems to deduce sensitive data from individuals' biometric information,[36] the use of real-time remote biometric identification systems in publicly accessible spaces for law enforcement purposes.[37] For other prohibited practices, oversight is shared between the DGCCRF and the French Regulatory Authority for Audiovisual and Digital Communication (ARCOM).

Therefore, the unification and coherence of AI regulation will primarily be ensured by the AI Office. While national authorities retain jurisdiction over high-risk AI systems within their respective borders, the AI Office holds exclusive competence for general-purpose AI systems. Additionally, it will oversee: cross-border high-risk AI systems, and high-risk AI systems built upon multiple general-purpose AI systems, to prevent jurisdictional overlaps with national authorities. In most cases, the AI Office will also intervene when national authorities encounter difficulties in obtaining sufficient information due to the involvement of cross-border operators, thereby limiting their national competence.[38]

It is therefore clear that administrative authorities overseeing data protection and artificial intelligence serve fundamentally different purposes. The former are primarily focused on safeguarding fundamental rights in the context of computer technology, while the latter are tasked with supervising the AI market as a product. However, reality is far more complex. This complexity arises from: the interconnected web of administrative authorities, the diverse range of competences, and regulatory overlap, particularly in cases where an AI system qualifies as personal data processing. In such scenarios, an AI operator may simultaneously be considered a data controller under the GDPR[39] and an AI operator[40] under the AI Act, creating dual regulatory obligations.[41]

This regulatory overlap, applied to the same subject matter, underscores the complementary yet distinct roles of these administrative authorities. Their broad and varied capabilities are particularly significant when fundamental rights protection is at stake. In this context, an assessment of how the French Data Protection Authority exercises its powers could provide valuable insights for anticipating the future enforcement of the AI Act.

2.2. The uncertainties surrounding sanctions by personal data authorities

The powers of supervisory authorities have evolved as regulations have adapted to the digital society we know today. These powers follow the same logic: the fundamental right of natural persons to the protection of personal data must be guaranteed, thanks to the specific role of supervisory authorities. Indeed, the supervisory authority acts both as a guide for controllers and as a judge in cases of regulatory violations. For example, the French Data Protection Act, which implements certain aspects of

[34] AI act, art. 5, §1, e).
[35] AI act, art. 5, §1, f).
[36] AI act, art. 5, §1, g).
[37] AI act, art. 5, §1, h).
[38] T. Douville, E. Netter, *The Artificial Intelligence Regulation: AI Law in Search of Coherence – Part 2*, in *RTD Com.*, 32, 2025; L. Badiane, M. Bourgeois, L. Bataille *et al.*, *AI Act: Authorities and Legal Remedies*, in *JCP E.*, 3, 2026.
[39] Art. 4, 7 GDPR.
[40] Art. 3, 8 AI Act.
[41] T. Douville, *Artificial Intelligence and Personal Data*, in *Dalloz IP/IT*, 2025, 147.

the GDPR, grants the authority the ability to "publish guidelines, recommendations, or frameworks intended to facilitate compliance with personal data processing";[42] to "handle complaints, petitions, and claims";[43] and, finally, to carry out inspections, either directly or through its staff, of any data processing operations and, where appropriate, to obtain copies of any documents or information media useful for the performance of its tasks.[44]

This consolidation of power must adhere to the principles set out since 1978 in Article 1, which stipulate that the development of information technology "must not infringe upon human identity, human rights, privacy, or individual and public freedoms". At the time, the legislator envisioned this administrative authority, the first of its kind, as "the guardian of societal awareness regarding the use of information technology".[45] However, from a procedural standpoint, this consolidation raised certain concerns. This is why the French Constitutional Council established a framework for this power when it approved the centralization of these prerogatives within administrative authorities,[46] on the condition that they are exercised by an independent body, that any proposed sanction excludes deprivation of liberty, and that the law provides measures to safeguard constitutional rights and freedoms.[47] These guarantees applied both to the French Data Protection Authority and to any other independent administrative authority. Consequently, the supervisory authority responsible for oversight was also required to follow the same principles.

Therefore, administrative authorities possess a range of powers to oversee the enforcement of both the GDPR and the AI Act. One key aspect of these powers is ex-ante supervision, which includes issuing guidelines and soft laws to assist controllers and AI operators in achieving compliance, as well as developing codes of practice.[48] For instance, in 2024, the French data protection authority responded to 1,448 requests for advice from controllers.[49] While this proactive approach is essential for helping controllers comply and enabling individuals to exercise their rights,[50] it is closely tied to the accountability principle which is a key aspect of the AI Act and GDPR.[51]

The accountability principle entails another set of powers, specifically the authority to sanction any breach of these regulations. To fulfill this mission effectively, administrative authorities possess extensive investigative powers to gather sufficient information about data processing. These powers include the right to access "any premises of the controller and the processor, including any data

[42] Art. 8, 2°, b, of French data protection act, mod. by the ord. n°2018-1125 du 12 déc. 2018 (French data protection act).
[43] Art. 8, 2°, d, French data protection act.
[44] Art. 8, 2°, g, French data protection act.
[45] B. TRICOT, *Report of the Commission Nationale de l'Informatique et des Libertés*, 1975, 89.
[46] Constitutional Council, Jan. 17, 1989, Law No. 88-248 DC amending the Sept. 30, 1986 law on freedom of communication.
[47] Constitutional Council, July 28, 1989, on the Law concerning financial market security and transparency, No. 89-260 DC.
[48] Art. 56, AI Act.
[49] CNIL, *Report of the Commission Nationale de l'Informatique et des Libertés,* 2025, 9.
[50] French data protection authority answer to 14 654 exercise of individual rights.
[51] A. LATIL, *Digital Law: A Risk-Based Approach*, 2023, 131; A. LATIL, *The Artificial Intelligence Regulation and the Risk-Based Approach: Application of a Structuring Legislative Method*, cit.

processing equipment and means".[52] In this context, controllers may respond to the report issued by the administrative authority or be heard by the restricted formation responsible for imposing sanctions.[53] In each of this step, the controller may recognize some of the facts, and none of the legal provision mentioned the right to remain silent. Therefore, a constitutional question has arisen regarding the scope of the right to remain silent, particularly in relation to the fundamental right against self-incrimination.[54] The French Supreme Administrative Court initially ruled that this right was not applicable during investigations conducted by the administrative authority. However, in response to a constitutional challenge, the French Constitutional Council determined that this right must be guaranteed.[55] Consequently, the provisions allowing controllers to submit observations in response to the administrative authority's report or to be heard by the restricted formation must be amended to incorporate this right. In the interim, the right to remain silent must be clearly defined. This incertitude abovementioned illustrate at the same time the extensive powers of French data protection authority, and some of the question remaining concerning the application of AI Act by administrative authority.

Despite the extensive powers granted to administrative authorities, every decision issued by the French data protection authority may be appealed before the Supreme Administrative Court (i.e. Conseil d'État). However, this procedural safeguard is undermined by the limited scope of judicial review exercised by the administrative court. The French data protection authority retains significant discretionary power over the outcomes of its investigations. For instance, it may "remind the party of its legal obligations or, if the observed breach is capable of being remedied, issue a formal notice requiring compliance within a specified timeframe".[56]

The authority enjoys broad discretion in determining how to address a complaint or an alleged breach of data protection regulations. The Supreme Administrative Court has consistently upheld this position,[57] even in cases where complaints are based on actual violations of personal data rights. This raises critical questions about the effectiveness of legal remedies available to individuals and the overall enforcement of data protection regulations.[58]

This broad discretionary power raises questions regarding another capability of the French Data Protection Authority. Since 2022, a simplified sanction procedure has been introduced to deal with cases that are not complex.[59] This procedure was established in response to the growing number of complaints. Several conditions must be met: first, other similar decisions must already have been issued by the French Data Protection Authority, and second, both the facts and the law of the case must be straightforward. The decision to follow this new procedure lies with the president of the Authority. The procedure is simpler because the president of the restricted committee decides alone. Finally, there is no hearing session for the case. Therefore, even though the controller may at any time request to switch

[52] Art. 58, point 1, f, GDPR.
[53] Art. 22, French data protection act.
[54] French declaration of Human and Civil rights of 26th august 1789, art. 9; Art. 6 ECHR.
[55]Constitutional Council, décision n° 2025-1154 QPC, 8 august 2025.
[56] Art. 20, II, French data act; Art. 58, point 2, GDPR.
[57] CE, 19 avr. 2024, n° 473459 ; CE, 10e et 9e ch. réunies, 21 oct. 2022, n° 459254.
[58] N. MARTIAL-BRAZ, CNIL, *Judicial Oversight and Data Governance: Between Discretionary Power and New Horizons*, in *Communication Commerce Electronique*, 2025.
[59] Art. 22-1, French data protection act.

to the ordinary procedure, financial sanctions under the simplified procedure are limited to €20,000, and may also take the form of a mere reminder or a compliance deadline. This simplified procedure may be a more effective way of handling the steady increase in complaints, but it lacks clear guidance on the use of sanctioning powers.

These uncertainty regarding sanction capabilities of the data protection authority raises question regarding AI regulation. The sanction procedure is not only intended to punish violations of data protection or AI regulations. It also has a preventive function, aiming to discourage others from acting in the same way. It provides regulations and soft law documents with another means of interpretation. However, this simplified procedure conceals the details of cases, the names of the parties, and the reasoning that led to the sanctions. These specificities, combined with the relatively low number of sanctions imposed by the French Data Protection Authority compared to its counterparts in Spain (281 cases, €35 million), Germany (416 cases, €14 million), and Italy (145 cases, €145 million),[60] highlight a more limited use of its sanctioning power. The French authority has issued only 87 sanctions, amounting to €55 million in total—including €50 million against Orange, a French telecom operator. Among these 87 sanctions, only 12 have been published. This means that there is no public information about the sanctioned controllers or the reasoning that led to the sanctions. The legal arguments are only briefly listed on the website of the French Data Protection Authority.[61]

While this procedure is a welcome development for handling multiple complaints, it does not fully serve the function of a sanction. To be truly effective, a sanction must fulfill a preventive function. Its purpose is indirectly to help other controllers achieve compliance by clarifying which practices constitute violations and by protecting individuals from being subjected to them. However, this new simplified procedure falls short of the true purpose of a sanction. It is undeniable that the significant fines imposed on GAFAM or BIATX send strong messages. Yet a violation of fundamental rights is of the same nature regardless of scale. The secrecy surrounding this simplified procedure largely benefits the controller while at the same time weakening the enforcement of the right to data protection.

These analyses are insightful regarding future application of the AI Act by administrative authority. Since French data protection authority will be in charge of any AIS that include personal data processing, the procedure and the jurisprudence in place will have an impact on how AI act will be enforce.

Regarding remedies, administrative enforcement is not the only means for individuals to seek redress or sanctions in the event of a GDPR or AI Act violation. Both regulations allow for the cumulative use of civil liability and administrative enforcement. Consequently, individuals may bring an action before the supervisory authority and also before a civil court. The latter option can serve as a way to compensate for any shortcomings in the administrative recourse.

3. The possible remedies available under civil liability

Civil liability law provides individuals with a means to seek compensation and to hold controllers accountable. This remedy is explicitly provided for in both the AI Act and the GDPR. Furthermore, these two legal avenues can be pursued simultaneously. Article 85 of the AI Act outlines the remedies

[60] EDPB, *Protection personal data in a changing landscape – EDPB Annual report*, 2024, 39.

[61] https://www.cnil.fr/fr/les-sanctions-prononcees-par-la-cnil (last visited 25/09/2025).

available to individuals or legal entities who believe that a breach of the regulation could warrant a sanction by the market supervisory authority. In such cases, the affected party may lodge a complaint with the designated administrative authority, but this is not their only option.

Indeed, this article mirrors the wording of Article 77 of the GDPR regarding remedies available to individuals. In essence, it states that "without prejudice to any other administrative or judicial remedy"[62] any person may file complaints with the independent administrative authority responsible for data protection or market supervision. Three avenues of recourse are therefore open to the individual: one before the civil courts, another before the administrative courts, and the last before the independent administrative authority. The text does not specify how these remedies interact. This is why a preliminary question was referred to the CJEU in the context of the exercise of the right of access to personal data by a data subject. The question was whether the individual could simultaneously file a complaint with the independent administrative authority and bring a civil action. In short, whether the remedies available under public law and those available under civil law could be exercised in parallel. The CJEU answered in the affirmative, stating that the remedies available to individuals under the GDPR may be exercised concurrently and independently.[63]

From this perspective, the reasoning applied by European judges under the GDPR can be extended to the remedies available under the Artificial Intelligence Act. The prohibition of certain practices under the AI Act would therefore not rely solely on administrative fines. Following the reasoning of the Court of Justice of the European Union outlined above, an individual who believes they have suffered harm as a result of such practices could both lodge a complaint with the relevant market surveillance authority and bring a claim before a judicial court.

The combination of these remedies is justified by two main considerations. First, they serve distinct functions: an administrative fine imposed by a national supervisory authority constitutes a public sanction targeting the AI system as a whole, while civil liability aims to compensate for individual harm. Second, the combination is justified by the regulatory framework governing AI systems, in which 'private enforcement' plays a central role.[64] This concept refers to the active participation of individuals in the enforcement of the regulation through civil actions.

However, unlike the GDPR, the AI Act does not allow for the personal liability of AI operators to be engaged. The AI Act does not provide an autonomous legal basis for civil liability claims. Instead, it follows the logic of treating AI as a product. As a result, the European Directive (EU) 2024/2853 on liability for defective products was amended to include AI-related damages. This means that individuals cannot seek compensation through the personal civil liability of the AI operator (A). One way to overcome these limitations could be to rely on civil liability under the GDPR to hold the AI operator

[62] For comparison, Article 77 of the GDPR states: "Without prejudice to any other administrative or judicial remedy, every data subject shall have the right to lodge a complaint with a supervisory authority (…) if he or she considers that the processing of personal data relating to him or her infringes this Regulation"; whereas Article 85 of the RIA provides: "Without prejudice to other administrative or judicial remedies, any natural or legal person who has grounds to consider that there has been a violation of the provisions of this Regulation may lodge complaints with the competent market supervisory authority".

[63] CJEU, 12 January 2023, Case C-132/21, *Budapesti Elektromos Művek*, ECLI:EU:C:2023:2.

[64] Private enforcement is a distinct notion from 'public enforcement', which refers to the action taken by public authorities to enforce a regulation: M.A. FRISON-ROCHE, J.C. RODA, *Droit de la concurrence*, 2022.

processing personal data accountable (B).

3.1. The limited scope of civil liability under the AI Act

Unlike the GDPR, the AI Act does not establish a separate legal basis for civil liability claims. This is primarily due to the abandonment of the proposed Directive on adapting non-contractual civil liability rules to artificial intelligence, which was discarded by the European Commission on 11 February 2025. The proposed Directive, as outlined by the European Parliamentary Research Service (EPRS), aimed to address "extra-contractual" civil liability—rules that would allow victims to seek compensation for harm caused by AI systems, regardless of any contractual relationship with the liable party. According to the EPRS, the Directive would have ensured that: "Any victim (individual or business) could be compensated if harmed by the fault or omission of an AI provider, developer, or user, resulting in damage recognized under national law (e.g., health, property, privacy, etc.)".[65] This form of liability was designed to be fault-based, meaning it would apply in cases where an AI operator breached a pre-existing legal obligation.[66] Combining this liability framework with the presumption mechanisms of the AI Act would have significantly strengthened protections for individuals harmed by AI systems. However, in a tense international political context,[67] the European Commission ultimately abandoned the Directive, citing simplification efforts as the primary justification.

Therefore, with the abandonment of this directive, the only remaining option for individuals is to invoke non-fault liability for defective products. This type of liability is almost exclusive, precluding other forms of liability. The Court of Justice of the European Union (CJEU) has limited the effects of the option provided by Article 13 of the directive.[68] This article allowed Member States a margin of appreciation in transposing the directive, meaning victims had to choose between liability based on a defective product or another common cause of liability. However, the CJEU restricted this choice to cases involving the producer's fault or warranty against latent defects, excluding the approach adopted by French civil judges based on the lack of expected safety.[69] As a result, the only way for victims to hold the producer liable is to prove damage separate from the defective product itself. For example, the French Cour de cassation (the supreme civil court) allows victims to act on the basis of a failure to provide information or a breach of the duty of care.[70]

In the context of AI, victims must prove that a defect in the AI system caused the damage.[71] This defect is defined as the system failing to provide the level of safety that a person is entitled to expect under EU

[65] M. TAMBIAMA, *Artificial intelligence liability directive*, in *Parliamentary Research Service*, 2023, 5.
[66] M. PLANIOL, *Elementary Treatise on Civil Law*, 11th ed., 1931.
[67] *EU Commission drops AI liability directive amid US criticism*, in *Harici*, 2025, disponible sur https://harici.com.tr/en/eu-commission-drops-ai-liability-directive-amid-us-criticism/, (last visited 15/09/2025).
[68] CJCE, 25 avril 2002, *Commission des communautés européennes c. République française*, C-52/00, D. 2002. 2462, chron. Larroumet; D. 2002. 1670, obs. Rondey; D. 2002. 2935, obs. Pizzio; CCC nov. 2002. Chron. 20, obs. Laporte; RTD civ. 2002. 523, obs. Jourdain; RTD civ. 2002. 868, obs. Raynard; RTD com. 2002. 585, obs. Luby.
[69] See, C. CAILLÉ, *Liability for Defective Products*, in *Rép. civ.*, 102, 2025.
[70] Civ. 1re, 7 mars 2006, 04-16.179, JCP 2006. I. 166, no 8, obs. Stoffel-Munck; RTD civ. 2006. 565, obs. Jourdain; RCA 2006, comm. 164, note Radé.
[71] On this matter: M. BACACHE, *Artificial Intelligence and the Law of Liability and Insurance*, in A. BENSAMOUN, G. LOISEAU eds. (dir.), *Droit de l'intelligence artificielle*, 2022.

or national law.[72] While the AI Act imposes significant obligations on AI operators and provides victims with multiple avenues to demonstrate such defects, operators may still invoke the 'development risk defense'. This exemption, as outlined in the relevant directives, allows operators to avoid liability if the damage was caused by a risk that could not reasonably have been discovered at the time the AI system was placed on the market.[73]

For products that, by their nature, evolve constantly after being placed on the market, this exemption significantly reduces victim protection under this liability regime. Furthermore, the victim's right to bring an action is subject to two statutory time limits. First, the victim has three years to act from the moment the damage occurs, the defect is discovered, and the identity of the AI operator is known.[74] Second, this special liability expires ten years after the product was placed on the market.[75] These time limits favor the producer (i.e., the AI operator) but restrict the victim's opportunities for redress, especially when compared to other liability regimes, such as fault-based liability, which allow a five-year period starting from the discovery of the damage.[76]

Then, the AI act does not provide direct ways for natural person to be compensate in case of AI violation. It's in this particular case where GDPR could be used by individuals to enforce AI Act provision as well as to be compensate in case of a damage linked with personal data violation.

3.2. The extent of remedies under the GDPR

As previously mentioned, the GDPR provides two types of remedies: a complaint before the administrative authority and civil liability of the controller in the event of damage caused by a GDPR violation. These two remedies can be pursued simultaneously for the same damage. However, these are not the only ways individuals can use the GDPR to seek compensation for violations of their personal data processed through an AI system.

Firstly, the right to data protection can be enforced through collective procedures against the controller. Data protection breaches are often collective issues, causing harm to multiple individuals. In such cases, the harm is considered "mass harm", defined as situations where "several individuals suffer individual injuries resulting from the same causative event".[77]

The French government, exercising the discretion granted to Member States, has empowered "a body, organisation, or not-for-profit association" to bring collective actions for effective judicial remedies against data controllers, particularly when the rights of individuals have been violated.[78] This mechanism represents a French adaptation of the US 'class action', but without punitive damages, only compensation for personal harm suffered by victims.

Several entities, including consumer associations and trade unions, are authorized to initiate such

[72] Directive (EU) 2024/2853 of the European Parliament and of the Council on liability for defective products and repealing Council Directive 85/374/EEC, 23 October 2024, art. 7, 1.
[73] *Ivi*, art. 11, point 1, e.
[74] *Ivi*, art. 16.
[75] *Ivi*, art. 17.
[76] Art. 2224 French Civil code.
[77] M. BACACHE, C. LARROUMET, *Obligations and Extra-Contractual Civil Liability: General Law and Special Regimesx*, 2021.
[78] GDPR, art. 80 §2.

actions. In practice, however, only two class actions were launched in 2019:[79] one by the Internet Society against Facebook, and another by the UFC-Que Choisir[80] against Google. There has been no public update on their progress since then.

Nonetheless, collective actions remain a potentially powerful tool for enforcing the GDPR, especially in light of the increasing number of data breaches in recent months, including several massive incidents.[81] They could serve as an effective complement to the actions (or inaction) of the CNIL, offering two key advantages for individuals: first, unlike administrative procedures, collective actions can directly award damages to victims; second, they reduce the burden on individuals, who might otherwise face time-consuming and costly legal proceedings.

Secondly, the Court of Justice of the European Union (CJEU) has adopted a broad interpretation regarding the actors who may base their legal action on a GDPR violation.[82] According to the CJEU, a breach of data protection regulations can be invoked not only by data subjects or entitled entities (such as consumer associations or labor unions), but also by other parties. This interpretation is grounded in Article 82 of the GDPR, which grants "any person" the right to seek compensation from the controller.[83]

This approach extends the scope of data protection regulation beyond the mere protection of the fundamental right to data protection. It also recognizes the economic value of personal data within the information society, where such data constitute a key competitive advantage for companies. After all, the GDPR pursues two main objectives: the protection of personal data and the free flow of data. The economic dimension of data is central to the digital economy. For this reason, the CJEU has allowed competitors to bring unfair competition claims based on GDPR violations, potentially leading to civil liability proceedings.[84]

These two avenues for engaging the civil liability of a data controller could serve as an interesting procedural remedy for the current lack of specific procedures addressing AI systems. Given that the processing of personal data is a central component of artificial intelligence systems, any violation of provisions related to personal data protection under the AI Act could potentially trigger the civil liability of the AI operator.

Indeed, the human-centric approach advocated by the European Commission places the responsibility for compliance squarely on the AI operator. Consequently, any operator of an AI system that processes personal data could be held liable for damages caused by such processing.

---

[79] https://observatoireactionsdegroupe.com/registre/registre-france/ (last visited 25/09/2025).

[80] A Consumer rights organisation.

[81] CNIL, *Report of the Commission Nationale de l'Informatique et des Libertés, 2024*, 505-629. Data breach have been listed by the french data protection authority, with some concerning millions of french individuals.

[82] CJUE, 4 octobre 2024, *Lindenapotheke*, C-21/23, D. 2024. 1777; *ivi*, 2115, point de vue F. Megerlin et E. Pinilla; Dalloz IP/IT 2025. 112, obs. V. Younès-Fellous; CCE 2024. Comm. 112, obs. A. Debet; CCC 2025. Comm. 7, obs. H. Aubry; RTD Com. 2025, p.94, note T. Douville; CJUE, 4 juillet 2023, *Meta Platforms e.a*, C-252/21, AJDA 2023. 1542, chron. P. Bonneville, C. Gänser et A. Iljic; D. 2023. 1313; Dalloz IP/IT 2024. 45, obs. A. Lecourt; RTD eur. 2023. 754, obs. L. Idot; CCE 2023. Comm. 94, N. Martial-Braz; Europe 2023. Comm. 340, obs. L. Idot; LEDICO, sept. 2023, DDC201u1, obs. T. Douville; CJUE, 28 avril 2022, *Meta Platforms Ireland Limited c/ Bundesverband der Verbraucherzentralen und Verbraucherverbände*, C-319/20, Dalloz IP/IT 2022, 461, obs. A. Latil; JA 2022, n° 660, 12, obs. X. Delpech; Dalloz IP/IT 2022. 229, obs. C. Crichton; RTD eur. 2023. 426, obs. F. Benoît-Rohmer; CCC 2022. Comm. 124, obs. S. Bernheim-Desvaux.

[83] CJUE, 4 juillet 2023, *Meta Platforms*, cit., nº 50.

[84] CJUE, 4 octobre 2024, *Lindenapotheke*, cit.

Furthermore, the core concept of damages related to the protection of personal data could take on a new dimension with the development of AI.[85] As this technology subtly or rapidly transforms every aspect of society, it may give rise to more nuanced and sensitive types of harm. This evolution could enable individuals to seek redress through the civil liability of AI operators and data controllers.

These two ways to engage civil liability of a data controller could be an interesting procedural palliative to the lack of special procedure for AIS. Since personal data processing are a key part of artificial intelligence systems, a violation of the provision that include personal data protection under AI Act could lead to engage civil liability of the AI operator. Indeed, the human centric approach wanted by the European commission tel the AI operator in charge of the compliance. Therefore, any operator of an AIS processing personal data could be responsible for any damages caused by it. More, the core concept of damages relating to protection of personal data could be gain a new dimension thanks to development of AI. Since this technology change subtly, or quickly, every aspect of our society, it could led to a more sensitive type of damages permitting individuals to act based on civil liability of AI operator and controller of personal data.

4. Conclusion

Far from reaching a conclusion, this section aims to synthesize the key issues at stake in the future development of AI regulation. As seen with the regulation of personal data, and the French perspective of this paper, the powers and role of independent supervisory authorities are central to safeguarding individual rights. However, the increasing complexity of these authorities as implemented in France, coupled with the scope of their powers, does not necessarily indicate a favorable trajectory for regulations that protect individual rights. For this reason, classic procedures derived from common civil rights should not be overlooked.

Despite the lack of specific civil action under the AI Act, the GDPR is likely to remain a common legal basis for civil actions against AI operators. Indeed, the CJEU has already expanded the scope of existing civil actions to allow the GDPR to be applied as broadly as possible (i.e allowing economic actor to act under GDPR for unfair competition). Consequently, this emerging and promising line of jurisprudence could be used to sanction AI operators that process personal data posing the greatest risks to individual rights and freedoms.

Yet, for the GDPR to effectively fulfill this role, it must remain unchanged. This is far from certain, given the proposal for an omnibus regulation that would amend both the GDPR and the AI Act, among others.[86] It will potentially weakened GDPR principles in order to facilitate AI system development.[87]

[85] On the notion of damages relating to personal data protection: J. KNETSCH, *The Compensation of Non-Pecuniary Loss in GDPR Infringement Cases*, in *JETL*, 13(2), 2022, 132.

[86] Proposal for a regulation amending Regulations (EU) 2016/679, (EU) 2018/1724, (EU) 2018/1725, (EU) 2023/2854 and Directives 2002/58/EC, (EU) 2022/2555 and (EU) 2022/2557 as regards the simplification of the digital legislative framework, and repealing Regulations (EU) 2018/1807, (EU) 2019/1150, (EU) 2022/868, and Directive (EU) 2019/1024 (Digital Omnibus), and amending Regulations (EU) 2024/1689 and (EU) 2018/1139 as regards the simplification of the implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI), 19 novembre 2025.

[87] Digital Omnibus Report V2: Analysis of Select GDPR and Privacy Proposals by the Commission.

Essays

Human Dignity and Quantified Self: The Constitutional Challenge of AI

Elena di Carpegna Brivio*

ABSTRACT: This essay examines the relevance of the constitutional concept of dignity in the digital society. This analysis examines how digital technologies are currently redefining the concept of human personality, employing a quantitative approach that considers human behavior through a statistical lens. The idea of dignity is then considered as a useful element for starting a new juridical reasoning aiming to draw a line of continuity through a person's physical, psychic, relational, and even digital existence, and the EU AI Act seems to be the beginning of a new regulation intended to define a technological development that could be authentically anthropocentric.

KEYWORDS: dignity; AI; fundamental rights; AI Act; quantified self

SUMMARY: 1. The Constitutional meaning of dignity and its relevance for the regulation of digital society – 2. The quantified essence of digital society – 3. Which dignity for the Quantified Self? The answer of the European Union – 4. The constitutional force of dignity for digital society.

1. The Constitutional meaning of dignity and its relevance for the regulation of digital society

The meaning of the principle of dignity is, even today, endowed with a certain degree of mystery, considering that it is difficult to find, in the field of Constitutional Law, another concept that is, at the same time, both central and elusive, a synthesis of the strongest juridical protection of human personality, yet also a source of numerous interpretative doubts. A partial explanation of this peculiarity could be found in the origin of the theoretical and philosophical elaboration of the concept, which is very ancient and predates the legal use of the word *dignity* by centuries.

In fact, the evolution of the meaning of dignity has characterized the entire development of Western civilization, with a moment of foundation at the beginning of Christianity, when the Church Fathers used the word to describe human beings as an image of God, an *Imago Dei*: off course, in that moment the reasoning about the concept was purely theological, but in that there was something grounding also from a juridical point of view, because, for the first time, through dignity, it was possible to consider human beings as universally equal because they all shared the same divine nature.[1]

During the humanist period, philosophers such as Pico della Mirandola and Pufendorf succeeded in

* *Tenure-Track researcher in Public Law. Mail: elena.dicarpegna@unimib.it.* *This article was subject to a blind peer review process.*

[1] P. RIDOLA, *La dignità dell'uomo e il 'principio libertà' nella cultura costituzionale europea*, in *Diritto comparato e diritto costituzionale europeo*, Torino, 2010, 84.

giving dignity an immanent connotation, strictly related to the rational nature of humans; however, the authentic turning point was Immanuel Kant's *Metaphysik der Sitten*. In the second part of his essay, Kant reinterpreted the concept's entire evolution, affirming that each human being has an inherent value grounded in the inner moral Law. Consequently, through dignity, Kant elaborated the idea that there is a system of rights that universally connotes every person.[2]

Through this path, the concept of dignity acquired its primary and current characteristics in Western legal culture.[3] It has become a principle capable of both expressing and grounding the idea that human rights are inherent to every individual, regardless of origin and social condition, but, despite this significant historical development, dignity was codified into Constitutions and the Charter of Rights only much later, because it was only after the atrocities of totalitarian regimes and the obliteration of the human person generated by nationalism, that dignity began to be written in Constitutions as intangible core of protection for the human person.[4]

The first legal text to explicitly refer to dignity is the Preamble to the Charter of the United Nations (1945). Then, in 1948, the Universal Declaration of Human Rights identified "the dignity of all members of the human family" as "the foundation of freedom, justice, and peace in the world" and established that "all human beings are born free and equal in dignity and rights". The authentic turning point, thus, was the new Constitutions of the two Countries, Germany and Italy, which had experienced the birth and rooting of totalitarianism.

The German Grundgesetz (1949) is undoubtedly the text that has had the greatest influence at the international level. In it, human dignity is the grounding principle of the whole constitutional system: Article 1 states that "Human dignity shall be inviolable. To respect and protect it shall be the duty of all State authority".

The Italian Constitution, although less well-known, has some significant peculiarities. Dated back in 1947, the Italian Constitution was the first that included dignity in its text; but, above all, the textual references to dignity (Articles 3, 36, and 41) define the concept as a limit that acts directly in economic-social relations and that must protect the human person not only from the power of the State but also from any private power. The result is an idea of a dignified life that binds all the juridical actors, public and private, and constantly combines rights and duties, freedom and solidarity.[5]

Through the achievements of those Constitutions, dignity became a new constitutional standard. After 1949, the concept of human dignity was incorporated into almost all European constitutions. Defining moments include the democratic transitions in Southern European countries (Greece, Portugal, and Spain) between 1975 and 1978, as well as the end of Communism in Central and Eastern Europe (1989-1991).[6]

However, the pinnacle of the entire process has been the European Union Charter of Fundamental Rights (2000). The Charter refers to dignity in the Preamble and affirms that "the Union is founded on the indivisible, universal values of human dignity, freedom, equality, and solidarity". Chapter I, then,

[2] I. KANT, *Die Metaphysik der Sitten*, 1785, part II.
[3] P. HÄBERLE, *Verfassungslehre als Kulturwissenschaft*, Berlin, 1998.
[4] M. ROSEN, *Dignity: its history and meaning*, Boston, 2012.
[5] A. RUGGERI, *Appunti per uno studio sulla dignità dell'uomo, secondo diritto costituzionale*, in *Rivista AIC*, 1, 2011, 16.
[6] C. DUPRÉ, *The Age of Dignity: Human Rights and Constitutionalism in Europe*, London, 2015, 55-56.

poignantly titled Dignity, articulates the meaning of the principle in the EU through a declaration of inviolability (article 1), the recognition of the right to life (article 2), of the right to the integrity of the person (article 3), the prohibition of the torture or inhuman or degrading treatment or punishment (article 4) and the prohibition of slavery and forced labour (article 5).

Through these codifications, dignity ceased to be mostly a 'never again pledge' to become an authentic grounding constitutional principle set by the Lisbon Treaty as the first fundamental value on which European integration is based (article 2).

The exact legal extent of the principle, however, has remained open. The scholars had defined dignity as a conceptual puzzle that can embrace at least four different interpretations: dignity as the prohibition of differences between human beings; dignity as the moral Law proper to each human being; dignity as a dignified way of acting; and finally, dignity as the right to be treated with dignity.[7]

Indeed, when examining the concrete application of dignity across various constitutional systems, the only certain data is the principle's jurisprudential operability. Nowadays, it isn't uncommon to find dignity invoked in decisions related to some debated issues like assisted suicide: in those kinds of decisions, dignity is the constitutional legal base for articulated reasonings about what can be perceived as a dignified way of living.

From a general perspective, it can be said that in courts, dignity serves as both an instrument to protect human liberty and a moral justification for public policies that safeguard specific values.

However, in recent years, a new field of application for human dignity has emerged, particularly in relation to the commodification of human personality that can be implied by the data economy and the integration of Artificial Intelligence into society.[8]

The next pages will analyze how human dignity has gained relevance in the legal debate over the regulation of Artificial Intelligence, especially following the approval of the EU AI Act in 2024.

The paper is articulated in three paragraphs. Paragraph two is dedicated to describing the quantitative essence of the digital revolution. It demonstrates how algorithms have increased the relevance of the mathematical-statistical perspective in society, with detrimental consequences for the idea that every person is unique and unrepeatable. The human person of the digital world isn't really a dignified human being, but a Quantified Self whose existence is shaped by statistical, actuarial, and probabilistic programs. That has several juridical consequences, considering how the quantitative mind is creating a new medievalism in which the feudal structure of the past is reproduced in the algorithmic classification of people. The paragraph will consider how digital architectures lead people to conform to the statistical normality of their category, thereby reducing opportunities for following unconventional paths and reinforcing prejudices and inequalities that constitutionalism has sought to eliminate for over two centuries.

Paragraph 3 examines then the legal instruments that the law seeks to establish to guarantee fundamental rights in a digital society, with a specific focus on the role the European Union has played since the implementation of the 2016 GDPR, which has affirmed that only a digital economy in which people participate with awareness and with a guarantee of their rights can become compatible with the

[7] S. CIVITARESE MATTEUCCI, G. REPETTO, *The expressive function of human dignity: a pragmatic approach to social rights claims*, in *European Journal of Social Security*, 2, 2021, 120.

[8] N. CASTREE, *Commodifying what nature?*, in *Progress in Human Geography*, 2003, 273.

idea of social progress (Recital 7). The paragraph considers the regulations approved by the EU following its 2019 digital strategy: the Digital Services Act, the Digital Markets Act, the Data Governance Act, the Data Act, and, finally, the Artificial Intelligence Act are the tools created by the European Union to shape all the juridical aspects of the digital society.
Paragraph 4, as a conclusion, highlights that the constitutional path of the dignity principle is now proving to be the primary legal tool for meeting the challenges of the technological future.

2. The quantified essence of digital society

Digital technologies are changing the way we live and work, and the more advanced the systems become, the more they are used to making relevant decisions: hiring, education, disease prevention, and even criminal convictions are just a few examples of decision-making processes in which the role of automated computing has become increasingly important.
In fact, a careful consideration of what is currently happening leads to reflecting on how the digital revolution could be considered as a transfiguration of human existence. As happened at the end of the 19th century with the Industrial Revolution, we are now experiencing a profound reinvention of everyday life that is reshaping social relationships, and at the core of this new era we can find a quantitative approach: to be understood by digital technologies, human life must be segmented and classified into elements detectable from a statistical and quantitative perspective.[9]
Consequently, digital technologies can perform optimally only when applied to contexts that are easily understood from a mathematical perspective, and the ideas grounding the constitutional development of fundamental rights cannot be readily translated into mathematical models. The juridical protection of human personality results from a complex juridical framework that relies on the idea that the limitation of power must be adapted to every possible situation to ensure a protection that should be the specific connotation of every individual.
The rising power of the digital society, thus, is quite different from every kind of power that Law has faced in the past: computational power has been meaningfully defined as a power that "is exercised by adapting or claiming to adapt, little by little, not only the world but also the representation of the reality to the functioning of digital information and communication technologies (ICTs)".[10] That means the digitalization works through segmentations that aim to place reality into a classification system designed to enable mathematical and statistical operations on the observed objects, and, so, only numerical elements are comprehensible in digitalization: similarities, differences, hierarchies, and correlations are the result of quantitative parameters present in the real world that can be perceived and reprocessed by algorithms.[11]
It is then clear that for the ICTs, social relationships assume the characteristics of statistical correlations: through the algorithmic filter, sociality is no longer based on people but on the quantitative elements

[9] L. Floridi, *The Fourth Revolution: How the Infosphere is Reshaping Human Reality,* Oxford, 2014.
[10] M. Durante, *Computational Power: The Impact of ICT on Law, Society and Knowledge*, 2021.
[11] E. Brynjolfsson, *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*, New York, 2014, 9; M. Broussard, *Artificial Unintelligence. How Computers Misunderstand the World*, Cambridge-London, 2018, chapter 3; T. Gillespie, *The relevance of algorithms*, in T. Gillespie *et al*. (eds.), *Media Technologies*, Cambridge-London, 2012, 167.

related to people's activities, which become the only tools for establishing affinities and differences, ordination and subordination, values and disvalues.

The digital revolution is, then, the result of applying increasingly powerful hardware and software to big data that, thanks to its volume, velocity, and variety, can easily reshape many elements of human existence, diminishing the relevance of the juridical framework that Law has defined to protect and promote fundamental rights and human personality.[12]

In the history of public Law, the moment the individual enters society has always been considered the moment of establishing relations of power. Only when humans become members of a community can they realize their personality through society.[13]

Consequently, juridical systems have consistently sought to establish specific models of juridical personality. It is well known how the French Revolution elaborated the structure of its legal system, starting from the idea of the *citoyen* as a being transfigured by Reason, a secular saint constantly dedicated to the common good and general interest.[14]

Similarly, after World War II, constitutionalism sought to rebuild a strong juridical protection of the person, emphasizing the idea of a human being connoted by his economic and social condition. That vision, as presented in the first paragraph, led to the anthropology of the *homo dignus*, meaning that every member of humanity should have fundamental rights that aren't mere claims regarding the State but are affirmed in society, using the force of the Law to assist a human existence that should be realized directly in intersubjective relations.[15]

Now, digital technologies affect the very assumptions on which constitutionalism has built the protection of the person, because they are not neutral facilitators of spontaneous relationships but rather architectures of affordances designed to guide behavior through selective, manipulative algorithmic pathways.[16]

The man of the digital society is a Quantified Self, a *homo numericus* increasingly engaged in producing data about itself, whether biological, physical, behavioral, environmental, or relational.[17] In the digital world, people are composed of multiple layers of data, and algorithms do not need any accuracy to operate on individuals, because the real power of digitalization lies not only in data collection but also in data aggregation. The strength of ICTs isn't knowledge about single elements but rather making large-scale comparisons and classifications that could infer a lot of information that, even if not accurate, are nonetheless effective.[18] When an algorithm succeeds in extracting those quantitative elements that enable a statistically significant aggregation or disaggregation, it possesses all the necessary information

[12] D. LANEY, *3D Data management: controlling data volume, velocity and variety*, META Group Research, 6, 2001.
[13] G. SARTORI, *Elementi di teoria politica*, Bologna, 1987, 241.
[14] G. BURDEAU, *La démocratie en chantier*, 1962, republished in G. BURDEAU, *Écrits de Droit constitutionnel et de Science politique*, Paris, 2011, 345.
[15] S. RODOTÀ, *Il diritto di avere diritti*, Bari, 2012.
[16] L. URQUHART, T. RODDEN, *New directions in information technology law: learning from human-computers interaction*, in *International Review of Law, Computers & Technology*, 2, 2017, 150.
[17] M. SWAN, *The Quantified Self. Fundamental Disruption in Big Data Science and Biological Discovery*, in *Big Data*, 2, 2013, 85.
[18] D.K. CITRON, F. PASQUALE, *The scored society: due process for automated predictions*, in *Washington Law Review*, 1, 2014, 1.

to perform.[19]

The Law, then, can no longer protect the individual simply by preventing interference with a private sphere, but must also prevent the construction of personality by those powers, public or private, who organize personal data for their own purposes.[20]

For example, algorithms often use data from a sample to infer conclusions about the entire population under study.[21] Similarly, they analyze past and present data to predict the future, or they can place people in groups that appear statistically similar but were created autonomously by the algorithms, maybe using randomization techniques.[22]

It then becomes tough to bring out the real merit and potential of a single individual, and instead, it's straightforward to encourage adaptation to statistical normality and the patterns identified by algorithms.

As a result, by calculating the probabilities of human behavior with increasing precision, it becomes much more difficult for the subject to choose alternative paths.[23]

In this context, the juridical condition of human personality can easily slide into a new medievalism. If algorithms determine which identities are standard and reject everything else, the uniqueness of each personality is gradually undermined by classifications that create a feudal system in which social layers are defined by statistical similarities and differences.[24]

The idea of a person as unique is replaced by the idea that everyone can be and do what is proper to the statistical category to which he or she belongs. By classifying a person, the algorithmic profile also defines the range of opportunities related to that person and, in doing so, brings subjects closer to the statistical normality of their categories.

Diversity is increasingly viewed as abnormality, and the most authentic core of constitutionalism is definitely compromised.

Human existence, considered for centuries as a core of self-determination embedded in given social relationships, is now becoming a statistical class membership.

To address this shaping power, it is necessary to guarantee the possibility of consolidating, as well as defining and communicating, a personality that is not hetero-determined, even in the face of continuous and pervasive algorithmic interferences.

It's in this direction that Law is trying to head at the moment, especially in Europe, where the European Union is attempting to define new rules to preserve the legacy of constitutionalism and the technological progress of humanity.

---

[19] J. VAN DIJCK, *Datafication, dataism and dataveillance: big data between scientific paradigm and ideology*, in *Surveillance and society*, 2, 2014.

[20] S. RODOTÀ, *Il diritto di avere diritti*, cit.

[21] M. HILDEBRANDT, *Learning as a machine: crossovers between humans and machines*, in *Journal of Learning Analytics*, 1, 2017, 6.

[22] J. KROLL *et al.*, *Accountable algorithms*, in *University of Pennsylvania Law Review*, 2017, 633.

[23] E. ESPOSITO, *The Future of Futures: The Time of Money in Financing and Society*, Cheltenham, 2011.

[24] P.P. VERBEEK, *Subject to technology*, in A. ROUVROY, M. HILDEBRANDT (eds.), *Law, Human Agency, and Autonomic Computing. The Philosophy of Law meets the Philosophy of Technology*, London, 2011, 27.

3. Which dignity for the Quantified Self? The answer of the European Union

When it comes to technology, the Law is fundamentally challenged. Compared to the dynamic development of digital technologies, the Law is a slow-moving entity.

In addition, the digital revolution operates with a degree of opacity that is difficult to match with the imperatives of transparency, certainty, and explicability required in a legal context.[25]

In recent years, legislators worldwide have sought to develop new strategies to safeguard fundamental rights in digital environments, and the regulation of artificial intelligence has become a new frontier for protecting individuals.[26] The approach of the European Union emerged as an attempt to replicate, in the field of Artificial Intelligence, the 'Brussels effect' that enabled the EU to present the regulation of privacy set out in the 2016 GDPR as a new global standard.[27]

The EU Commission outlined, in the communication COM(2021)118, a comprehensive digital strategy up to 2030 that has, as its main goal, to shape the EU as an advanced digital marketplace, in which the stereotype of a Law that chases technological innovation is replaced by a digital economy in which people enjoy strong guarantees of personal rights.

In this document, the Commission outlined an approach completely different from the past: the European regulation was no longer conceived as a legislation mainly oriented to the market, but has taken on the characteristics of what it seems to be a proper constitutional discipline, with a primary focus on the strengthening of the adequate protection of fundamental rights as enshrined in the Charter of Fundamental Rights of the European Union.[28]

As discussed in the first paragraph, the Charter is built on the juridical concept of dignity, and that made the European Digital Strategy strongly committed to defining for the future a digital environment firmly rooted in the prevention of any kind of exploitation of human beings through the accumulation, aggregation, and reuse of personal data.[29]

The EU Directive 2019/770, as the first implementation of the Strategy, openly addressed the risk of commodification of human personality that could be associated with the economic value of personal data.[30]

Then, in a few years, five regulations (the Digital Services Act, Digital Markets Act, Data Act, Data Governance Act, and AI Act) rapidly reshaped the European approach to digitalization.

The Digital Services Act (Regulation UE 2022/2065) has updated the rules on the liability of platforms in

---

[25] H. RUSCHEMEIER, *AI as a challenge for legal regulation – the scope of application of the artificial intelligence act proposal*, in *ERA forum*, 2023, 361.

[26] A significant relevance should be recognized for some institutional analyses developed in recent years, such as the UK AI national strategy, presented to Parliament by the Secretary of State for Digital, Culture, Media and Sport in September 2021. Another cognitive institutional insight has emerged from hearings held by the Subcommittee on Privacy, Technology, and Law of the USA Senate under the Biden Administration.

[27] A. BRADFORD, *The Brussels Effect: How the European Union Rules the World*, Oxford, 2020.

[28] G. PITRUZZELLA, *Big Data, Competition and Privacy: a Look from the Antitrust Perspective*, in *Concorrenza e Mercato*, 2016, 15 ff.; F. COSTA-CABRAL, O. LYNSKEY, *Family Ties: The Intersection Between Data Protection and Competition in EU Law*, in *Common Market Law Review*, 2017, 11 ff.

[29] A.C. WITT, *The Digital Markets Act – Regulating the Wild West*, in *Common Market Law Review*, 60, 2023, 625 ff.

[30] Directive (EU) 2019/770, *on certain aspects concerning contracts for the supply of digital content and digital services*, 20 May 2019. Significantly, e.g., the Recital 24, which explicitly declares that "the protection of personal data is a fundamental right and that therefore personal data cannot be considered as a commodity".

providing digital services.

The premise of DSA is the increasing dominance of digital platforms in managing vast amounts of personal data. However, the regulation also includes thoughtful reasoning on the potential impact of such disposal of personal data on an individual's personality. The Digital Services Act, in fact, imposes several obligations on platform operators, differentiated by the size and volume of data they manage. There are four types of intermediaries: brokerage services, hosting services, online platforms, and large-scale platforms. All of them are obliged to introduce a reporting and action mechanism (Art. 16), which must allow any subject to notify the platform of the presence of illegal content. Article 17 then requires platforms to explain any restriction of visibility or removal of content. At the same time, Articles 20 and 21 require operators to have a complaints management service, and Article 21 provides extrajudicial dispute-resolution mechanisms that should be easily accessible via the online interface.[31]

The main innovation, however, lies in highlighting the systemic risks inherent in the activities of large-scale platforms and research engines. These risks are identified in the dissemination, through digital services, of illegal content or in the possibility of negative effects on the fundamental rights enshrined in the Nice Charter, and, mainly, on human dignity, private and family life, protection of freedom of expression and information, non-discrimination, respect for the rights of the child, and protection of consumer rights.[32]

Additionally, the Digital Markets Act (DMA), Regulation (EU) 2022/1925, established a new framework for the digital economy. In fact, the DMA opened a new era in which Article 102 TFEU assumes a particular meaning for some large digital companies, known as gatekeepers, which are subject to antitrust obligations and prohibitions that apply preventively.

According to Article 3, gatekeepers are companies that have a significant impact on the internal market, provide a core platform service which is an essential gateway for business users to reach end users, and enjoy an entrenched and durable position in its operations, or it is foreseeable that it will enjoy such a position shortly.[33]

Article 5, then, prescribes that gatekeepers must seek users' consent to combine their data across different services and must provide an equivalent alternative to users who decline consent.

The premise of the European reasoning is that the size of these entities implies a relevant and unavoidable influence that can select the market options available to the individual, and ultimately, can nudge people into giving consent for data treatment only because an authentic alternative isn't provided. Therefore, the EU imposed specific obligations: the measures include the interoperability of services, surveillance of data generated on platforms and in advertisements, the elimination of lock-in mechanisms for promoting bids and concluding contracts, equal treatment of services provided by other gatekeepers, allowing consumers to connect with companies outside the platform, and the uninstallation of any pre-installed software or apps. Additionally, the European Commission assumed direct supervision and control over the application of the DMA, thereby superseding the previous

[31] C. PINELLI, *L'evoluzione della normativa dell'Unione europea*, in C. PINELLI, U. RUFFOLO, *I diritti nelle piattaforme*, Torino, 2023, 13 ff.

[32] A. TURILLAZZI, M. TADDEO, L. FLORIDI, F. CASOLARI, *The digital services act: an analysis of its ethical, legal and social implications*, in *Law, Innovation and Technology*, 1, 2023, 83.

[33] Regulation (EU) 2022/1925, *Digital Markets Act*, article 3.

competences of national authorities, and initiated various surveillance procedures for the six Big Tech companies identified as gatekeepers (Alphabet, Amazon, Apple, ByteDance, Meta, and Microsoft).[34]
Using that approach, the EU has made clear that it's no longer possible to affirm the principle of the neutrality of digital platforms. What is now emerging is a clear awareness that digital technologies are powerful instruments and, as such, can produce subjection, discrimination, and inequalities that should be strongly prevented by the Law.
However, the pinnacle of the entire European ruling process is the Artificial Intelligence Act, proposed by the Commission in April 2021 and definitively published in July 2024 as Regulation (EU) 2024/1689. In it, respect for human dignity is emerging as a significant legal principle that informs the AI's ruling.
As is well known, the main feature of the AI Act is to address the complex world of Artificial Intelligence through a risk-based approach. In fact, the Regulation identifies different risk levels, corresponding to various regulation scenarios: unacceptable risk technologies, high-risk technologies, limited-risk technologies, and low-risk technologies. Unacceptable risk technologies are prohibited by the EU, high-risk technologies are subject to strict rules and obligations, and limited-risk technologies must meet transparency obligations. Low-risk technologies can be freely placed on the EU market.
The choice of a risk-based approach is effective because it allows for a comprehensive ruling action, avoiding the pitfalls of the notion of AI, which is not only ambiguous but also rendered uncertain by ongoing developments in the field. Using the risk-based approach, the Regulation can potentially cover any data processing technology; however, the provision's vagueness is balanced by the tangible social impact that can be ascribed to a single technology.[35]
The Act addresses algorithmic classifications in various parts of its discipline.
In Chapter II, dedicated to prohibited AI practices, the EU has considered the commercialization of systems that can persuade people into unwanted behaviors, or that nudge them into decisions in a way that subverts and impairs their autonomy and free will, to be an unacceptable risk to fundamental rights (Recital 29).
Article 5, then, lists the practices that can produce this kind of effect, considering it unacceptable to deploy subliminal techniques or to exploit the vulnerability of a natural person or a group of people due to their age, disability, or a specific social or economic situation. The letter c) specifically prohibits the commercialization of AI products that could evaluate or classify natural persons or groups of persons over a certain period based on their social behavior or known, inferred, or predicted personal or personality characteristics. The provision, however, limits its applicability to situations in which the algorithmic social scoring could be detrimental to individuals or groups of individuals in a social context unrelated to the context in which the data was originally collected, or could produce an effect that is disproportionate to the observed social behavior and its gravity.
The article also prohibits systems that could assess or predict the risk of a natural person committing a criminal offense based solely on the profiling of a natural person or on assessing their personality traits and characteristics (letter d); create or expand facial recognition databases through the untargeted scraping of facial images from the internet or CCTV footage (letter f); infer emotions of a natural person in the areas of workplace and education institutions (letter g); use of biometric categorization systems

[34] The identification was made on the 6th of September 2023.
[35] J. SCHUETT, *Defining the scope of AI regulations*, in *Law, Innovation and Technology*, 1, 2022, 1.

that categorize individually natural persons based on their biometric data to deduce or infer their race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation (letter i).

Additionally, the regulation of high-risk technologies focuses on mitigating the severe effects of algorithmic classifications. According to the provisions of Annex III, the areas in which AI technologies can develop some high-risk effects are: biometrics, with a specific attention to remote biometric identification/categorization or emotion recognition; education and vocational training; employment, workers management and access to self-employment; access to and enjoyment of essential private services and essential public services and benefits, with a specific reference to systems intended to be used to evaluate the creditworthiness of natural persons or establish their credit score; law enforcement; migration, asylum and border control management; administration of justice and democratic processes.

As can be seen, the AI Act pays significant attention to the possibility that artificial intelligence could have discriminatory effects, thanks to its quantitative approach, rooted in its statistical nature. In the discipline, there's a specific intent to prevent some of the main distortions realized by artificial intelligence in the first years of its development.

Nonetheless, this remarkable achievement is in a text that originally sought to balance the protection of personal rights with the EU's interest in strengthening the economy of personal data. It is important to remember that the initial proposal for the AI Act, made by the Commission in April 2021, was considering the protection of personal rights as a question of the lack of trust that individuals can develop in AI technologies (Whereas 1 and 5 of the text proposed by the EU Commission). However, the European Parliament's discussion of the proposal helped to change the approach. The rapid advancement of generative artificial intelligence has demonstrated how easily AI can generate complex texts, images, videos, and audio that can significantly alter human perception, opening the field to a vast range of new threats to human rights.

This circumstance prompted the EU to partially rethink the structure of the AI Act, including in the Regulation a set of general principles that must ensure that Artificial Intelligence is developed with full respect for the human person. Dignity is closely tied to the new framework incorporated into the final text.

Specifically, following parliamentary amendments in July 2023, a new general goal was added to Article 1, defining the primary purpose of the AI Act as ensuring a human-centric, trustworthy artificial intelligence that always respects the fundamental rights enshrined in the EU Charter. In addition, the final text of the Regulation now provides some limits that apply to all types of AI, regardless of risk level, and should help define an artificial intelligence that is, by design, ethical, reliable, and fully compliant with the Union's Charter of Fundamental Rights.

The boundaries of AI legitimacy encompass human intervention and surveillance, technical robustness and safety, respect for privacy and data governance, transparency, respect for diversity, nondiscrimination, fairness, and social and environmental welfare.

The final result is a far cry from the original balance between person and market, and it appears to clearly prioritize protecting human rights over the function of any technology.

The AI Act now explicitly acknowledges that a regulation on AI should ensure that technology is always

serving people and that people consistently exercise their rights.
Textual references to human dignity are prevalent throughout the AI Act, and their global significance lies in reconstituting the multiple dimensions that the human person assumes in digital environments.
As a result, dignity is emerging as the element that can draw a line of continuity through a person's physical, psychic, relational, and even digital existence.[36]

4. The constitutional force of dignity for digital society

The connection between dignity and the AI Act enables consideration of fundamental rights in a manner that differs significantly from the usual development of constitutional systems.
In fact, under European law, the old bilateral pattern of fundamental rights is no longer current. In a digital society, personal rights can't be exercised solely within relationships with public powers: people define their identities in a network of relationships that are both real and digital, and data flows are increasingly relevant in the construction of social perceptions and representations. State power isn't the primary threat to individual freedom, and the strategy of liberating people solely by protecting their privacy from outside interference is no longer effective.
Digital technologies can influence the human mind even before the process of willingness can begin, and every digital behavior can be prompted by a statistical model that assigns the highest probability to the desired result. That means that human personality is constantly exposed to systemic risks that must be addressed not only by public powers, but especially by private actors who, having hegemony in data management, are consequently also developing hegemony over individuals and their rights.[37]
Dignity has been introduced in Constitutions precisely as an element that prevents challenging the rights and freedoms that outline the anthropological texture of constitutional and democratic orders. Dignity has the juridical ability to simultaneously limit and direct social organization, and as such, it could also be the main juridical criterion for a constitutional turn in the liquid nature of digital society.[38]
Therefore, through dignity, it becomes possible to give a constitutional interpretation to digital society: if globalization has promised the removal of barriers and has actually built deep interconnections and interdependencies, it does not follow that the human condition should be transfigured according to the characteristics of computational power.[39]
As seen in the previous pages, the juridical concept of dignity emerged from the most tragic moments of European history. Dignity is written in the Constitutions to introduce an undecidable element, a core idea that prevents the principles, rights, and freedoms that define the anthropological texture of democratic systems from being challenged.[40]
Its inner force in contemporary society, however, isn't just about closing the door on an unwanted past.

[36] S. RODOTÀ, *Il diritto di avere diritti*, cit.
[37] R. MESSINETTI, *La tutela della persona umana versus l'Intelligenza Artificiale. Potere decisionale dell'apparato tecnologico e diritto alla spiegazione della decisione automatizzata*, in *Contratto e impresa*, 2, 2019, 885.
[38] Z. BAUMAN, D. LYON, *Liquid Surveillance: A Conversation*, Cambridge-Malden, 2013.
[39] R. BROWNSWORD, *What the World Needs Now: Techno-Regulation, Human Rights and Human Dignity*, in R. BROWNSWORD (ed.), *Global Governance and the Quest for Justice*, Portland, 2004, 203 ff.
[40] S. RODOTÀ, *Antropologia dell'homo dignus*, in *Storia e memoria*, 2, 2018, 107 ff.; H. ARENDT, *The Origins of Totalitarianism*, London, 1966, 3.

Its actual specific ability is to anchor the system of subjective rights to an objective principle that, as such, is subtracted from any balancing technique. As an evocative synthesis of universal rights, dignity cannot be rediscussed and must remain immutable, even in the era of exponential technological change.[41]

In a world where digital technology is deeply integrated into society and can easily influence the definition and development of the Self, the relevance of dignity lies in imposing respect for the autonomy of the individual on all social actors, public powers, and private individuals. It is the principle that reminds everyone that the Law must always ensure the free and full development of the human person.[42]

It is possible to affirm that the legal limitation of Artificial Intelligence cannot in any way be based on simple mechanisms of referral to human self-determination and must instead follow multi-principled approaches that make the legal protection of the person a structural element of algorithmic architectures.[43]

Developing a moral and social obligation in respecting the unique value of every individual, dignity expresses the legal duty to prevent any use of technology that could alter the equal relevance of every person.[44]

The presence of dignity in almost every Constitution of European Countries and in the EU Charter of Fundamental Rights can be regarded as an important starting point, awaiting to be played out in all its consequences and potential for application. Dignity is the juridical principle that is now strongly affirming the legal duty to prevent the subjection of human beings to any predetermined pattern. It commits the Law to developing social actions capable of ensuring that each person's social existence is the result of everyone's ability to contribute, according to one's own possibilities and choices, to the spiritual and material progress of society.[45]

The presence in Europe of a strong network of Constitutional Courts and National Data Protection Authorities provides the legal infrastructure that could make dignity a sound principle, preserving for future citizens the heritage of more than two Centuries of constitutionalism.[46]

[41] E. DI CARPEGNA BRIVIO, *Pari dignità sociale e Reputation scoring. Per una lettura costituzionale della società digitale*, Torino, 2024.
[42] L. FLORIDI, *The Fourth Revolution*, cit.
[43] L. FLORIDI, *Toleration and the Design of Norms*, in *Science and Engineering Ethics*, 21, 2015, 1095 ff.
[44] Q. CAMERLENGO, *Costituzione e promozione sociale*, Bologna, 2013.
[45] S. RODOTÀ, *Il diritto di avere diritti*, cit.
[46] B. MARCHETTI, *The algorithmic administrative decision and the human in the loop*, in *Biolaw journal*, 2, 2021.

Essays

Artificial Intelligence and Media Freedom: From the 'Brussels' to the 'Strasbourg' Effect?

Giovanni Zaccaroni*

ABSTRACT: Artificial intelligence (AI) poses both challenges and opportunities for media freedom in Europe. The interplay between EU legislation – including the Media Freedom Act, the AI Act, and the Regulation on political advertising – and the European Charter and ECHR seeks to protect media pluralism and democracy amid rapid digital transformation. AI's growing role in content creation and distribution raises concerns over autonomy and editorial independence, but EU law aims to ensure transparency, AI literacy, and safeguards against undue influence. Cooperation between the EU and the Council of Europe through frameworks like the AI Convention reinforces a safe approach to innovation, promoting digital strategic autonomy for the EU

KEYWORDS: artificial intelligence; media freedom; Council of Europe; European Union; ECHR

SUMMARY: 1. AI and media freedom in the EU approach to the digital sector – 2. Media freedom and AI technologies in the media industry – 3. The EU legal framework on media freedom and AI – 4. Towards a 'Strasbourg' effect? – 5. Conclusion: the case for joining forces.

1. AI and media freedom in the EU approach to the digital sector

This paper will outline how the regulation of AI under EU law, not solely in the AI Act, can allow the emergence of an EU approach to digital strategic autonomy. The media industry and its high reliance on AI is an example where primary and secondary sources of EU law interact and shape an approach to regulation that is crucial for EU digital strategic autonomy. Such an approach can be strengthened by the interaction between the EU and the Council of Europe framework, in particular on AI.

The relationship between media freedom, pluralism, and AI emerged in particular in recent years.[1] The importance of a media industry that is free and plural is increasingly considered as a fundamental characteristic of a democratic society.[2] Artificial intelligence, on the other hand, is increasingly used in the

* *Associate Professor of EU law, Università degli Studi di Milano-Bicocca. Mail. giovanni.zaccaroni@unimib.it.* *This article was subject to a blind peer review process.*

[1] A. MUNORIYARWA, M.F. DE-LIMA-SANTOS, *Generative AI and the Future of News: Examining AI's Agency, Power, and Authority*, in *Journalism Practice*, 19, 2025, 2177.

[2] R. MASTROIANNI, *Freedom of pluralism of the media: an European value waiting to be discovered?,* in *Media Laws*, 2022, 100. C. SCHEPISI (a cura di), *Unione europea, pluralismo e libertà dei media nell'era digitale,* Napoli, 2025.

media industry to generate and distribute content, and the automatization of this process poses important questions to the freedom and pluralism of media.[3]

This paper aims to address the following research question: how, in the example of media freedom and pluralism, AI regulation can shape an approach that is distinctive of the EU, and whether this approach, through the interaction and cooperation between the EU and the Council of Europe, can contribute to the consolidation of EU digital strategic autonomy.

To address the question, it will be necessary to discuss the application of primary legislation, as well as the use of the legal basis (mostly art. 114 TFEU) to adopt several legal instruments that are still within a transitory regime. Such a regime, in the coming years, will increase uncertainty on the application of EU legislation and of the legal framework in the field of AI and media freedom. On the other hand, it will be important to discuss the application of AI in the media industry, as in examples such as in media content distribution, in fact-checking, and in content moderation, and its impact on media freedom. The application of EU legislation, both at primary and secondary levels, also happens in a field where, for obvious reasons, there is very little *specific* case law.[4] Such a case law falls outside the scope of this work, as at present this analysis is mainly focused on the primary and secondary sources of EU legislation.

In general, EU law requires that the life cycle of AI-based products and services should embed an approach that protects and promotes media freedom and media pluralism, by providing tools that can enhance the respect of this pluralism.[5] At the same time, the tools should not be so restrictive that they constitute a permanent barrier to the development of an innovative media industry. The development of an AI that reaches this balance will eventually contribute to the EU's digital strategic autonomy. To reach this goal, it is important to assess the potential role of the Council of Europe in promoting the EU approach to AI (as long as such an approach is compatible with the one of the Council of Europe). There might be mutual benefits, but also mutual challenges, in the cross-fertilisation between two largely overlapping but distinct European legal orders.

2. Media freedom and AI technologies in the media industry

The media industry is rapidly changing, especially after the advent of the World Wide Web and the transition to the digital sphere.[6] The main source of income of media outlets is no longer (with few exceptions) the daily purchase but rather the alternative between a monthly (or yearly) subscription and advertising.[7] The need to attract investors that are willing to spend resources on media outlets is forcing

[3] *Guidelines on the responsible implementation of artificial intelligence systems in journalism*, 30 April 2025, https://rm.coe.int (last visited 31/01/2026).

[4] However, there is an impressive body of *general* case law, both from the Court of Justice of the EU and from the European Court of Human Rights on media freedom, that can help and guide the interpretation of these very new pieces of EU legislation. See D. VOORHOOF, T. MCGONAGLE, *Freedom of Expression, the Media and Journalists: Case-law of the European Court of Human Rights*, 2017, https://rm.coe.int (last visited 31/01/2026).

[5] See in general, T. BLAGOJEV *et al*., *Monitoring Media Pluralism in the European Union,* 2025, https://cadmus.eui.eu (last visited 31/01/2026). At p. 21, the Report also explicitly mention the risks for the use of AI in journalism.

[6] J.P. SIMON, M. BOGDANOWICZ, *The Digital Shift in the Media and Content Industries,* in *JRC Scientific and Policy Reports,* 2012.

[7] E. BROGI, H. SJØVAAG, *Good practices for sustainable news media financing,* 2024, https://rm.coe.int (last visited 31/01/2026).

media companies to change their business models to produce a larger amount of content that should be, at the same time, shared and promoted not only on the web but into a variety of social media outlets, which requires, in turn, a variety of different approaches.[8] Also, the traditional media industry is challenged by the competition of freelance media personalities (e.g., individuals whose YouTube channel or Instagram account has more than several hundred thousand, if not several millions, of subscribers).[9] This does not mention the fact that traditional and more authoritative media outlets have considerably more expenses compared to freelance media personalities, which makes their traditional business model even less sustainable.[10]

The need to produce more content that can be shared on multiple platforms to compete with multiple actors itself justifies the reliance on AI, which allows media outlets to automatise production of content and to tailor it to the need to the different platforms.[11] To this, it should be added that AI technology is provided on a pay-per-use basis, which makes it difficult for media outlets to develop proprietary AI systems as the costs are a magnitude higher than the cost of a subscription.[12]

By joining the elements of the high competition, the increase in the costs, in the number of contents to be produced and in the number of platforms where the content is to be shared, it can be understood the degree of potential dependency of media industry and of journalists on AI.[13] This dependence has been defined in scholarship as "infrastructural reliance" as the media platform are unwilling, and perhaps legitimately so, to build and develop their own AI infrastructures, but they rely intensively on the one provided by other actors, as Microsoft, Google and OpenAI.[14]

Such a relationship can, in turn, be an important factor of innovation, but can also affect the freedom and the pluralism of the media industry.[15]

---

[8] J. HENDRICKX, J. VÁZQUEZ-HERRERO, *Dissecting Social Media Journalism: A Comparative Study Across Platforms, Outlets and Countries,* in *Journalism Studies*, 25, 2024, 1053.

[9] E.g. MrBeast, one of the freelance media personalities with more subscribers on YouTube, has 464 million subscribers, while the YouTube account of the Financial Times has 1.39 million. If we restrict our search to Italian media industry, Luis Sal (renowned Italian YouTube personality) has 1.88 million subscribers on YouTube, while *Il Corriere della Sera* (the most reputable and diffused traditional Italian newspaper) has 487k subscribers (as of 31 January 2026).

[10] M. SAVAGE, *Social media creators to overtake traditional media in ad revenue this year* in *The Guardian,* 10 June 2025, theguardian.com (last visited 31/01/2026).

[11] A. NANZ, A. BINDER, J. MATTHES, *AI in the Newsroom: Does the Public Trust Automated Journalism and Will They Pay for It?,* in *Journalism Studies*, 2025, 1.

[12] *Ivi*, 6-7.

[13] T. HOLLANEK, D. PETERS, E. DRAGE *et al., AI, journalism, and critical AI literacy: exploring journalists' perspectives on AI and responsible reporting*, in *AI & Society*, 2025. F.M. SIMON, *Escape me if you can: How AI reshapes news organisations' dependency on platform companies* in *Digital Journalism*, 2024, 149-170. F.M. SIMON, *Uneasy bedfellows: AI in the news, platform companies and the issue of journalistic autonomy,* in *Digital journalism*, 2022, 1832. N. HELBERGER, M. VAN DRUNEN, S. ESKENS, M. BASTIAN, J. MOELLER, *A freedom of expression perspective on AI in the media – with a special focus on editorial decision making on social media platforms and in the news media* in European Journal of Law and Technology, 2020, ejlt.org (last visited 30/01/2026). P. PARCU, M.A. ROSSI, *Policy changes to strengthen the protection of media freedom and media pluralism in the EU,* in P. PARCU, E. BROGI (eds) *Research Handbook on EU Media Law and Policy,* 2020, 427.

[14] M.Z. VAN DRUNEN, *Safeguarding media freedom from infrastructural reliance on AI companies: The role of EU law,* in *Telecommunication Policy,* 49*,* 2025.

[15] *Ivi*, 6.

The media industry, and especially journalists, play an important role in a democratic society, which is one of promoting and sharing content that is reliable and authoritative.[16] The monopoly, or rather oligopoly, on the sources of production of content is giving the upper hand to AI companies and social media platforms to potentially restrict and influence access to the main sources of media content distribution. A recent example, in this sense, is the refusal by Meta Platforms to accept requests for political advertising following the entrance into force of EU legislation in the field.[17]

The use of AI technologies in the distribution and sharing of content is an undeniable opportunity, but it is also becoming, with time, an important challenge. It is thus important to understand to what extent this relatively new interaction is regulated by EU law and also how the EU regulatory approach can contribute to its own strategic autonomy.

3. The EU legal framework on media freedom and AI

In the following section, a recognitive analysis of the regulation applicable to media freedom and pluralism and to AI in the EU legal framework is hosted. As this area is developing certainly at a high pace, the reader should be aware of the fact that the framework is changing at an unprecedented pace.

3.1. EU primary law

The Treaties themselves do not focus on media freedom, and there is no explicit legal basis that authorises the EU institutions to act to protect it. There is, however, an interesting recognition of the value provided by public funding to the media industry that is relevant to this analysis and is contained in Protocol no 29 to the Lisbon Treaty "on the system of public media broadcasting".

This Protocol is interesting as, since its inception, it defines "that the system of public broadcasting in the Member States is directly related to the democratic, social and cultural needs of each society and to the need to preserve media pluralism". This consolidates the assumption that media freedom and pluralism are a precondition for democracy within the EU legal order.[18]

This Protocol adds that the provisions of the Treaties, and in particular the ones on State aid and on public procurement, should be understood as not preventing the Member States "from providing funding to public service broadcasting and insofar as such funding is granted to broadcasting organisations for the fulfilment of the public service remit […]".

What can be understood from a teleological interpretation of this Protocol is that the pursuit of the objectives of the EU, including the ones about the single market, can be limited to protect the media broadcasting services that are fulfilling a public service within the Member States.

---

[16] T. Dodds, R. Zamith, S.C. Lewis, *The AI turn in journalism: Disruption, adaptation, and democratic futures,* in *Journalism*, 2025, journals.sagepub.com (last visited 31/01/2026).

[17] *Ending Political, Electoral and Social Issue Advertising in the EU in Response to Incoming European Regulation,* 25 July 2025, about.fb.com (last visited 31/01/2026).

[18] See in this sense also the initiative promoted by Reporters Sans Frontières under the name of *Paris Charter on AI and Journalism,* 10 November 2023, https://rsf.org (last visited 31/01/2026). On AI and democracy see O. Pollicino, P. Dunn*, Intelligenza artificiale e democrazia: opportunità e rischi di disinformazione e discriminazione*, Milano, 2024.

This should not be considered as an exact rule, but rather, in particular in light of the legal value of Protocols, an approximation that can help in the interpretation of the EU Treaties and of primary legislation.[19]

3.2. EU fundamental right instruments

While the Treaties offer limited guidance on the matter, the same cannot be said for fundamental rights instruments that either play the role of primary sources of EU law or that find their way to display legal effects through other provisions of the Treaties.

The Charter of Fundamental Rights of the European Union acknowledges the respect of 'freedom and pluralism of media' within its Art. 11, paragraph 2, on freedom of expression and information.[20] This second paragraph, according to the Charter Explanations, extends paragraph 1 to media freedom and pluralism and is based on the case law on television of the Court of Justice.[21]

In the *Stichting Collectieve* case (expressly recalled in the Explanations of the Charter), the Court of Justice already in the early '90s maintains that "The maintenance of the pluralism […] is connected with freedom of expression, as protected by Article 10 of the European Convention on Human Rights and Fundamental Freedoms, which is one of the fundamental rights guaranteed by the Community legal order".[22]

Now the exercise of media freedom and pluralism focuses considerably less on television and much more on the digital sphere, although this will take some time to reach the EU courts.[23] A recent example of case law (2024) on media freedom, the *Real Madrid* case, is about the limitations and the financial sanction imposed for the publication of an article that appeared in the newspaper *Le Monde* in December 2007.[24] In that judgement, the Court of Justice maintained that "Article 11 of the Charter constitutes one of the essential foundations of a pluralist, democratic society, and is one of the values on which, under Article 2 TEU, the European Union is founded".[25]

Article 10 of the European Convention of Human Rights, although not an EU instrument, also has a content that is very similar to that of art. 11 of the Charter and has nurtured a case law of the European

[19] On the legal value of Protocols see e.g. L. PECH, *The European Union's Lisbon Treaty: Some thoughts on the 'Irish Legal Guarantees',* in *EJIL Talk!,* 28 September 2008, ejiltalk.org (last visited 31/01/2026).

[20] Art. 11, *Charter of Fundamental Rights of the European Union*: "1. Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers. 2. The freedom and pluralism of the media shall be respected".

[21] Judgment of the Court of 25 July 1991, case C-288/89, *Stichting Collectieve Antennevoorziening Gouda and others v Commissariaat voor de Media*.

[22] Case C-288/89, para 23.

[23] On recent case law on media freedom see R. MASTROIANNI, *I nuovi confini della libertà di informazione nel diritto dell'Unione europea,* in *Rivista AIC,* 2025, 233.

[24] Judgment of the Court (Grand Chamber) of 4 October 2024, case C-633/22, *Real Madrid Club de Fútbol and AE v EE and Société Éditrice du Monde SA.*

[25] Case C-633/22, para 49.

Court of Human Rights that is also founded on the role of media pluralism and freedom within the EU society.[26]

There is however another, non-binding, fundamental rights instrument that is particularly interesting for our analysis, and it is the Declaration on Digital Rights and Principles for the Digital Decade.[27] This soft-law instrument translates the vision of the EU for digital rights and represents an important update to the existing body of legislation on media freedom and pluralism, as it refers explicitly to new phenomena as artificial intelligence and the divide between the 'online' and the 'offline' environment.[28]

This Declaration explicitly mentions, in its Chapter IV, the participation in the public digital sphere as one of the characteristics of the EU approach to the digital environment. In particular, it states that access to multiple sources of information represents a key to a healthy online environment: "Access to diverse content contributes to a pluralistic public debate and effective participation in democracy in a non-discriminatory manner".[29] The Declaration also explicitly mentions the role of very large online platforms in granting support to the democratic debate online, as well as safeguarding fundamental rights online, and in particular media freedom and pluralism.[30]

The framework of the fundamental rights instrument, both binding and non-binding, explicitly recognising the role and the importance of media freedom and pluralism in the digital environment and its interaction with AI, is thus particularly lively. There are, however, several challenges, in particular in terms of enforcement at the national level, that this primary framework fails to address and that might be better implemented at the level of secondary legislation.

3.3. EU secondary law

Multiple sources of secondary EU law capture the intersection between AI, media freedom, and pluralism. These acts are largely adopted under the same legal basis, Art. 114 TFEU, either alone or in conjunction with Art. 16 TEU.[31]

---

[26] Art. 10.1, *European Convention of Human Rights*: "Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers. This Article shall not prevent States from requiring the licensing of broadcasting, television or cinema enterprises".

[27] *European Declaration on Digital Rights and Principles for the Digital Decade*.

[28] A. ADINOLFI, *Evoluzione tecnologica e tutela dei diritti fondamentali: qualche considerazione sulle attuali strategie normative dell'Unione*, in *Quaderni Aisdue*, 2023, 321. E. CELESTE, *Digital Constitutionalism, EU Digital Sovereignty Ambitions and the Role of the European Declaration on Digital Rights*, in *New Directions in Digitalisation*, 2025, 255.

[29] *European Declaration on Digital Rights and Principles for the Digital Decade*, para 12.

[30] *European Declaration on Digital Rights and Principles for the Digital Decade*, para 15: "Online platforms, particularly very large online platforms, should support free democratic debate online. Given the role of their services in shaping public opinion and discourse, very large online platforms should mitigate the risks stemming from the functioning and use of their services, including in relation to misinformation and disinformation campaigns, and protect freedom of expression. We commit to: a) continuing safeguarding all fundamental rights online, notably the freedom of expression and information, including media freedom and pluralism […]".

[31] On the use of legal basis see E. LONGO, *Grounding Media Freedom in the EU: The Legal Basis of the EMFA* in *Rivista italiana di informatica e diritto*, 7, 2025, 14. P. DE PASQUALE, *Dalla flessibilità delle basi giuridiche alla normazione integrata: tecniche legislative funzionali alla rigidità del riparto di competenze nell'UE*, in *Il Diritto dell'Unione europea*, 1, 2025, 1.

The first one is the Media Freedom Act (MFA), whose main provisions recently became directly applicable,[32] that provides a general framework at the EU level to make sure that media services provided within the EU digital single market are delivered in a way that preserves media freedom and protects democracy.[33]

The key provisions of the MFA turn around Art. 4, on the rights of media service providers, and Art. 6, on the duties.[34] Art. 4, in particular, being on the rights of media service providers, translates into obligations for the Member States to protect and promote media freedom at the national level.[35] A little bit further in the text of the MFA (that is not as conspicuous as the AI Act), we find Art. 18, which is about the functionalities that providers of very large online platforms (VLOPs) should grant to their users. In this provision is explicitly provided that VLOPs should not include content generated with artificial intelligence without previously disclosing to the users.[36] This rule reinforces the explicit obligation for online platforms to disclose all the contents that are AI-generated in other EU legal acts.[37] However, the provision goes even further, as it obliges the VLOPs not to display AI content that has not been subjected to human review or editorial control. It is unclear to what degree the VLOPs have already started to comply with this obligation, as the majority of them mainly apply, so far, transparency obligations.

Article 22 of the MFA is also important for our analysis, as it provides for the creation, at the national level, of rules to avoid media concentrations.[38] These rules should also allow the EU and the MSs to monitor the media industry and to potentially understand when reliance on certain technology, including AI infrastructure, is producing a potential distortion of competition in the relevant market.

The AI Act, although more general, is also an important piece of legislation for AI and media freedom.[39] The risks posed by an AI system to media freedom should surely be taken into account not only in the risk management system under Art. 9 of the Regulation, but also for the fundamental rights impact assessment in Art. 27 of the same Regulation (both for high-risk AI system ex Art. 6 of the Regulation).[40] However, for the object of this paper is perhaps more important the sometimes neglected Article 4, on AI literacy. This Article provides that the "AI systems shall take measures to ensure, to their best extent, a sufficient level of AI literacy of their staff and other persons dealing with the operation and use of AI systems on their behalf". This is not a banal statement, seen from a media freedom perspective. The AI Act is, in fact, introducing an obligation for media service providers to make sure that their staff are

[32] As of 8 August 2025, with certain exceptions in Art. 29 of Regulation (EU) 2024/1083.

[33] Regulation (EU) 2024/1083 of the European Parliament and of the Council of 11 April 2024 establishing a common framework for media services in the internal market and amending Directive 2010/13/EU.

[34] M.C. ORISTANO, *L'Unione europea e la tutela dei diritti fondamentali nel settore dei media alla luce dello European Media Freedom Act,* in *Eurojus*, 4, 2025, 16.

[35] On Art. 4 obligations towards EU Member States see L. MALFERRARI, *New and reinforced rights for media service providers under Article 4 European Media Freedom Act*, in *Rivista Italiana di Informatica e Diritto,* 7, 2025, 1.

[36] Art. 18.1 e), Regulation (EU) 2024/1083.

[37] E.g. Art. 50 of Regulation (EU) 2024/1689.

[38] Art. 22.1, Regulation (EU) 2024/1083: "1. Member States shall lay down, in national law, substantive and procedural rules which allow for an assessment of media market concentrations that could have a significant impact on media pluralism and editorial independence".

[39] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence.

[40] On the fundamental rights impact assessment see: A. COSENTINI, O. POLLICINO, G. DE GREGORIO *et al.*, *Assessing the Impact of Artificial Intelligence Systems on Fundamental Rights*, 2025, https://ssrn.com (last visited 31/01/2026).

trained in the use of AI and that they are able to recognise and evaluate the risks that are posed by their use.[41] AI literacy is, in fact, one of the most important actions that can be taken, still at a relatively contained cost, especially compared to that of a potential lawsuit.

The last piece of legislation to be analysed is the Regulation on the transparency of political advertising.[42] This legislation has the ambition to fight against the external influence in the realm of national and EU elections, and to disclose the source of funding of the political and social advertising that are showed also in the media industry, including online.[43]

This Regulation, which became directly applicable very recently,[44] mentions fundamental rights, and in particular, media freedom and pluralism, since its inception.[45] In an attempt to substantiate this very ambitious commitment, the Regulation also establishes rules for the targeting and ad-delivery techniques in online political advertising.[46] These provisions explicitly restrict the collection of personal data for political advertising, and establish also transparency obligations that should disclose how these targeting and ad-delivery techniques work.[47] These transparency obligations should, in particular, disclose to what extent an artificial intelligence system has been used in the targeting or ad-delivery technique.[48]

This Regulation proves to be very important for the EU's strategic autonomy, since, in a situation where they can only limit control of the foreign influence on the digital sphere, it is crucial to have transparency on the sources of funding and on the use of AI.

However, the online media platforms did not stand still and refused to comply with the obligations in the Regulation. Meta Platforms, in particular,[49] decided to openly oppose this Regulation since its negotiation phase, and since June 2025, they have openly renounced political advertising in the EU territory, a decision recently enforced.[50]

This behaviour undoubtedly represents a defeat for the EU approach to regulation in the field of AI and media freedom. In this way, the group that manages, among others, Facebook and WhatsApp has hinted that they are ultimately not interested in an important source of funding coming from the EU. At the same time, they suggest that they can afford not to comply with the EU legislation, and that, for them, this non-compliance is even to a certain extent convenient. It is perhaps not by chance that Meta Plat-

---

[41] T. HOLLANEK, D. PETERS, E. DRAGE *et al., AI, journalism, and critical AI literacy: exploring journalists' perspectives on AI and responsible reporting*, cit., 3.

[42] Regulation (EU) 2024/900 of the European Parliament and of the Council of 13 March 2024 on the transparency and targeting of political advertising.

[43] L. LIONELLO, *La reazione europea alle interferenze nei processi elettorali: il nuovo regolamento relativo alla trasparenza e al targeting della pubblicità politica*, in C. SCHEPISI (a cura di), *Unione europea, pluralismo e libertà dei media nell'era digitale,* Napoli, 2025, 277.

[44] As of 10 October 2025, *ex* Art. 30 of Regulation (EU) 2024/900.

[45] Recital (65), Regulation (EU) 2024/900: "When complying with their obligations under this Regulation, providers of political advertising services should pay due regard to fundamental rights, and other rights and legitimate interests. Providers of political advertising services should in particular pay due regard to freedom of expression and information, including media freedom and pluralism".

[46] Art. 18, Regulation (EU) 2024/900.

[47] Art. 19, Regulation (EU) 2024/900.

[48] Art. 19.1 c), Regulation (EU) 2024/900.

[49] However, similar concerns have been raised by Microsoft (who stopped political advertising in the EU since 2019) and Google.

[50] *Backlash as new EU political ad rules kick in,* 10 October 2025, https://www.politico.eu (last visited 31/01/2026).

form is the online platform that has the most advanced system of self-regulation of content moderation, which resembles that resembles the one of a national judiciary, chaired by its Oversight Board.[51]
This move by Meta Platform can, however, also be interpreted as being on the right track for the EU: a defensive reaction from one of the largest online platforms suggests the unwillingness to disclose the sources of political advertising or to perform such a collection.
There is, however, a similar development related to another proposal of the European Commission that should also be considered here.
In 2022, the European Commission proposed a Regulation on preventing and fighting sexual abuses, which is currently being negotiated by the Parliament and the Council.[52] This Proposal, in its initial form, provides for the establishment of a Coordinating Authority in each of the EU Member States to fight against child abuse.[53] These Authorities have the power to request that the national judicial authorities intervene to detect and remove the online content that can be connected with child abuse.[54] This proposal, however, has been understood in several EU Member States (and in particular in Germany) as threatening the privacy of the citizens: this has triggered an enormous e-mail campaign that has eventually convinced some EU Member States to vote against the Regulation, a fact that most likely will make its adoption.[55] This last example, although not directly related to the use of AI in the media industry, has very similar consequences and implications to the Meta Platform *affair*: an EU legislation that pursues a legitimate aim (fighting child abuse) gets misunderstood at the political level, and it becomes untenable. Ultimately, either the proposed piece of legislation is abandoned, or it becomes obsolete.[56]
Considered altogether, the lesson for the EU is that if there is a need to promote its own approach, it cannot be only achieved through hard law, but mostly through negotiation and moral suasion. In this case, the example of soft law in other areas of EU law can pave the way to a more balanced approach.[57]

4. Towards a 'Strasbourg' effect?

The trajectory of this paper started with the assumption that, in light of the role that media freedom and pluralism undertake within EU democracy, and in light of the importance that AI has in the online

[51] D. WONG, L. FLORIDI, *Meta's oversight board: A review and critical assessment,* in *Minds and Machines,* 2023, 261.
[52] Proposal for a Regulation of the European Parliament and of the Council laying down rules to prevent and combat child sexual abuse, COM (2022)209.
[53] Art. 25, Proposal for a Regulation of the European Parliament and of the Council laying down rules to prevent and combat child sexual abuse.
[54] Art. 7, Proposal for a Regulation of the European Parliament and of the Council laying down rules to prevent and combat child sexual abuse.
[55] *One-man spam campaign ravages EU 'Chat Control' bill,* 8 October 2025, https://www.politico.eu (last visited 31/01/2026).
[56] Something similar happened with the proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive) COM (2022) 496 final. It was perceived as overlapping with the AI Act and the Commission was accused of hyper regulation. In early 2025 the Commission announced the intention to abandon the proposal. The proposal on the digital euro is also currently subjected to this fate.
[57] M. ELIANTONIO, E. KORKEA-AHO, U. MÖRTH (eds), *Research Handbook on Soft Law*, 2022, Cheltenham. P.L. LÁNCOS, N. XANTHOULIS, L. ARROYO JIMÉNEZ (eds) *The Legal Effects of EU Soft Law*, Cheltenham, 2023. M. ELIANTONIO, A. VOLPATO, S. RÖTTGER-WIRTZ (eds), *Global Standards and EU Law*, Cheltenham, 2025.

media industry, the realm of media freedom is a compelling example of how the EU's approach to AI regulation can foster its own strategic autonomy in the digital realm. It has also emerged, in particular in the last part of Section 3, that an EU approach that is based solely on hard law encounters limitations in the ambition of large online media platforms to self-regulate themselves, as well as in the initiative of civil society and in the public debate.[58] The accusations moved to the EU institutions are accordingly two main ones: hyper-regulation and excessive intrusion in the private dimension.

These accusations are very different. While the latter is purely political, and as such, the bigger role is played by the democratic debate within the EU institutions, the first one is legal in nature and can be dealt with by regulatory instruments. The claim that the EU is hyper-regulating the technological sector might be widely exaggerated, but it has some truth in it.[59]

In order to address this claim, the EU needs to take seriously the requests that come to simplify the regulatory framework, but it is also equally, if not more important, to engage in an activity to mainstream its own regulatory approach within other legal orders. The EU has been an example in economic integration around the globe: certain CARICOM Member States recently decided to implement free movement of persons in a move that mimics the EU internal market.[60] It can equally lead the way in the approach to technology. This is not only about the so-called 'Brussels effect', that is a solid and serious approach that, however, relies largely on unilateral decisions, but through a multilateral effort that not only involves the EU, which can perhaps be called the 'Strasbourg effect'. It is in Strasbourg in fact, that the Council of Europe is located.[61] The Council of Europe has been studying AI since the mid 2010 and, since 2022, has been working on a Convention on AI and Human Rights, Democracy and the Rule of Law.[62]

This Convention has been opened for signature on 9 September 2024, and it has been signed by the European Commission, on behalf of the EU, on that very day.[63] At present, together with the EU, which has signed the Convention on behalf of its Member States, 16 more States that are either Contracting Parties or Observers of the Council of Europe have also signed.[64] Some of these States are candidate coun-

[58] I. NENADIĆ, R. CARLINI, O. SPASSOV, *A decade of digital transformation: Pluralism between the media and digital platforms*, in E. BROGI, I. NENADIĆ, P. PARCU (eds) *Media Pluralism in the Digital Era: Legal, Economic, Social, and Political Lessons Learnt from Europe*, Routledge, 2024, 17, 25-26.

[59] The number of legal acts proposed or adopted since 2015 under the legal basis of Art. 114 TFEU (alone or in conjuction with other Articles) is unprecedented, and spans from data protection to cryptoassets and digital euro to artificial intelligence.

[60] *Barbados, Belize, Dominica and St. Vincent and the Grenadines Ready for Full Free Movement on 1 October 2025,* 30 September 2025, Caricom.org (lastly accessed 12/10/2025).

[61] The Council of Europe is famous for having promoted the *European Convention of Human Rights* and its court, the European Court of Human Rights. However, it has so far been engaged in the promotion of more than 200 Conventions on a variety of subjects.

[62] *Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*, Council of Europe Treaty Series n. 225.

[63] 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.*

[64] So far, Andorra, Georgia, Iceland, Liechtenstein, Montenegro, Norway, Republic of Moldova, San Marino, Switzerland, Ukraine, United Kingdom, Canada, European Union, Israel, Japan, United States of America, Uruguay have signed the AI Framework Convention.

tries to enter the EU, and might have joined the Convention under these auspices.[65] Others, however, are not EU Member States and might not necessarily associate the EU approach to AI.[66] Although under the old administration, even the United States signed the AI Framework Convention, and so far, the new administration has not taken explicit action against it.

The AI Framework Convention is an international instrument that, according to authoritative scholarship, can not be exactly superimposed with the EU legislation in place.[67] However, there are legal innovations that allow for this Convention to become a vehicle to contribute to mainstreaming an approach to AI that can help the EU to enhance its strategic autonomy.

The AI Framework Convention has two main legal innovations: one is Chapter III on "Principles related to activities within the lifecycle of artificial intelligence systems",[68] and the other is Chapter V on "Assessment and mitigation of risks and adverse impacts".[69] Among the principles related to the life cycle of AI systems, it should be mentioned Article 13, which is the new "Principle of safe innovation".[70] Chapter V, on the other hand, deals with the two legal tools that characterise the EU approach to AI in the AI Act and that are summarised under a single article, Article 16 on "Risk and impact management framework".[71]

An interpretation that can be suggested in light of the reading of Article 16 is that the instrument brings together the risk management system and the fundamental rights impact assessment under the AI Act. This approach is further developed into a version that the Committee on Artificial Intelligence, which promoted the Convention, calls "risk and impact assessment of artificial intelligence (AI) systems from the point of view of human rights, democracy and the rule of law (HUDERIA)".[72] In any case, it is undeniable that Article 16 shares many common points with at least the content of art. 27 of the AI Act, and it is known that the AI Framework Convention has been negotiated during the same period as the AI Act, an indication of conceptual convergence and likely mutual influence.[73]

Despite of course these similarities, the EU legislation on AI and the AI Framework Convention are different legal instruments. However, they are adopted under a similar approach, that is, the one that the innovation developed in the EU should be safe and should not undermine democracy, fundamental rights, and the rule of law.

---

[65] E.g. Georgia, Montenegro, Moldova and Ukraine.

[66] E.g. Canada, Switzerland, United Kingdom, Israel.

[67] J. ZILLER, *The Council of Europe Framework Convention on Artificial Intelligence vs. the EU Regulation: two quite different legal instruments,* in *CERIDAP*, 2, 2024, 202.

[68] Arts. 6 to 13, *AI Framework Convention*.

[69] Art. 16, *AI Framework Convention*.

[70] The substance of the Article is about the possibility, for State parties, to use sandboxes as regulatory experiments. See para 60, Explanatory Report to the *Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*.

[71] Para 105, Explanatory Report to the *Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*.

[72] See Committee on Artificial Intelligence, *Methodology for the Risk and Impact Assessment of Artificial Intelligence Systems from the Point of View of Human Rights, Democracy and the Rule of Law*, 28 November 2024, www.coe.int/ (last visited 12 October 2025).

[73] See also recital (5) of Council Decision (EU) 2024/2218: "[…] the personal and material scope of the Convention and the substantive provisions of the Convention coincide to a large extent with Regulation (EU) 2024/1689 […]".

This common approach, despite some differences, is the one that should convince the EU to ratify the Convention.

The AI Framework Convention is, at the present stage, to be ratified by the States that have already signed it, and it will enter into force after the ratification of 5 signatories, at least three of which should be Council of Europe Member States.[74] The AI Framework Convention is an international agreement that is necessary, for the EU, to attain the objective of harmonising national legislation in the field of AI.[75] International agreements in the field of competences that are shared between the EU and its Member States[76] can be concluded as 'EU only' agreements or as 'mixed' agreements, concluded by the EU and Member States.[77] If the agreement is concluded as 'EU only', the EU will exercise the powers deriving from membership in the AI Framework Convention, including the power to vote in the Conference of the Parties,[78] on behalf of all of the Member States.[79] If one or more of the EU Member States decides to ratify autonomously the AI Framework Convention, then it will trigger the possibility that all of the other EU Member States will have in turn to ratify it, slowing considerably down the ratification process, if not making it impossible.[80] To maximise the effectiveness of the AI Framework Convention for the EU and its Member States, thus crucial to ratify the AI Framework Convention as soon as possible as an 'EU only' agreement, following the procedure for international agreements in art. 218 of the TFEU.[81] This ratification will, in turn, make it more likely that the 'Brussels' effect (the EU approach) is mainstreamed in Strasbourg by the Council of Europe.

5. Conclusion: the case for joining forces

In the introduction of this paper, it was claimed that the EU's approach to the regulation of the use of AI in the media industry has radical implications for media freedom and pluralism. These implications are expressed in the attempt to limit and to control, as much as possible, the relationship of dependency that media platforms develop towards the owners of AI infrastructures, as well as in the automation of key aspects of their own daily work. At the same time, the use of AI in the media industry can contribute to reducing the costs and increasing the opportunities, hence very often the pluralism, of the media space. Ultimately, a media industry that develops a healthy relationship with AI infrastructures is neces-

[74] Art. 30, *AI Framework Convention*.

[75] See recitals (4) and (5), Council Decision (EU) 2022/2349 of 21 November 2022 authorising the opening of negotiations on behalf of the European Union for a *Council of Europe Convention on artificial intelligence, human rights, democracy and the rule of law*.

[76] In the case of the *AI Framework Convention*, it is extremely difficult to determine if the EU has exclusive competence or not. The Council decision on the opening of the negotiations itself says that it should be negotiated as an EU only agreement but also contemplates the possibility that other Member States will subsequently ratify it. See recitals (6) and (9), Decision (EU) 2022/2349.

[77] P. CONCONI, C. HERGHELEGIU, L. PUCCIO, *EU Trade Agreements: To Mix or Not to Mix, That Is the Question*, in *Journal of World Trade*, 55, 2021, 231.

[78] Art. 23, *AI Framework Convention*.

[79] Art. 216.2 *TFEU*: "Agreements concluded by the Union are binding upon the institutions of the Union and on its Member States".

[80] P. CONCONI, C. HERGHELEGIU, L. PUCCIO, *op.cit.*, 242.

[81] J. HELISKOSKI, *The procedural law of international agreements: A thematic journey through Article 218 TFEU*, in *Common Market Law Review*, 57, 2020, 79.

sary for the development of EU democracy and for the democratic space of the EU Member States. The promotion and the protection of EU democracy is equally a fundamental part of the EU strategy to advance its autonomy and independence in the digital sector.

However, in order to foster this approach to EU digital strategic autonomy, the EU lawmakers should join the binding regulatory effort with the ability to persuade the media industry, and in particular online platforms, to comply autonomously with its own regulatory framework. This is possible only if the EU approach, which is characterised by the idea that innovation should not happen at the cost of fundamental rights and democracy, consolidates at the international level as the alternative to an approach that promotes innovation at all costs. This approach can be successful only if other legal orders choose to embrace and support this approach. This is what is happening, with, of course, certain differences, at the level of the relationship between the EU and the Council of Europe.

That is why it seems important for the EU and its Member States to ratify the AI Framework Convention. By actively mainstreaming the EU approach through the AI Framework Convention, it becomes for the EU a little less difficult to promote its own regulatory approach and to resist the temptations to abandon an approach based on safe innovation, which many large online platforms (like Meta and X) are at present actively opposing.

Essays

Workplace Neurosurveillance: Is the Employee's Mental Privacy Protected Under International Law?

Marta Sosa Navarro*

ABSTRACT: This paper analyses the implications for international human rights and labour law resulting from the use of neurotechnology in the workplace. It distinguishes between brain-reading devices, which collect and process neural data and may affect privacy and freedom of thought, and brain-altering technologies, which may affect mental integrity. By mapping the international, regional, and ILO frameworks, this paper highlights protection gaps created by fragmented regulation of this disruptive technology. It argues that the precautionary principle, soft-law instruments, and anticipatory regulation are essential to address these challenges. Ultimately, it contends that safeguarding dignity in the digital workplace requires a principled and proactive governance model to prevent cognitive surveillance and exploitation.

KEYWORDS: neurotechnology; workplace surveillance; mental privacy; freedom of thought; international labour law

SUMMARY: 1. Introduction – 2. International human rights system of protection against workplace surveillance – 2.1. Privacy as a precondition: special focus on Convention 108+ of the Council of Europe and Article 17 ICCPR – 2.2. The right to freedom of thought in the workplace under international human rights law – 3. The International Labour Organization and emerging challenges of workplace surveillance – 3.1. ILO safeguards of mental privacy – 3.2. ILO safeguards of mental integrity – 4. An overview of the human-centered but fragmented European normative framework protecting the worker's mental privacy and mental integrity – 4.1. The role of Article 88 GDPR in workplace personal data processing – 4.2. The EU Medical Device Regulation 2017/745 – 4.3. The AI Act and the regulation of emotion-inference technologies in the workplace – 5. Conclusive remarks.

1. Introduction

In a 2022 report commissioned by the British Law Society on the implications of neurotechnology for law and the legal profession, Australian criminal law scholar Allan McCay drew attention to the risks of workplace neurotechnologies by introducing the provocative idea of a "billable unit of attention".[1] In this report, McCay cautioned that the collection and processing of neural data through attention-monitoring neurotechnologies could allow law firms to adopt forms of neurosurveillance disguised as billing innovations, charging clients for the "measurable attention"

* *Assistant Professor in International Law, University of Milano-Bicocca. Mail: martamaria.sosanavarro@unimib.it.*
This article was subject to a blind peer review process.

[1] A. MCCAY, *Neurotechnology, law and the legal profession*, in *Horizon Report for The Law Society on Neurotechnology*, 2022, 26.

devoted to their case rather than for hours worked. Although still a hypothetical scenario, current market and investment trends in consumer neurotechnologies, coupled with the post-COVID expansion of workplace surveillance, suggest the emergence of a trend likely to penetrate everyday life, often unnoticed behind the smokescreen created by the contemporary normalization of self-tracking practices, which require individuals to disclose highly personal information, including health and biometric data, to private corporations.[2] These dynamic underscores the urgency of assessing whether, in a context of increasingly invasive technology-driven workplace monitoring, the international human rights system is adequately equipped to safeguard workers' rights and dignity.

Neurotechnology refers currently to devices, systems, and procedures — encompassing both hardware and software — that directly measure, access, monitor, analyze, predict or modulate the nervous system to understand, influence, restore, or anticipate its structure, activity and function.[3] If we consider that this AI-powered technology, already embedded in consumer wearables such as earbuds or headbands,[4] can measure neural activity,[5] infer cognitive, emotional, and neurological states, connect the brain to digital systems via brain-computer interfaces (BCIs),[6] and alter brain functioning,[7] its transformative impact beyond the individual, extending to society as a whole is not hard to imagine.[8]

As neurotechnologies continue to advance and move beyond clinical contexts, where they have significantly improved the lives of patients with neurological conditions such as Parkinson's disease, epilepsy, locked-in syndrome, and treatment-resistant depression[9] into direct-to-consumer markets, international debate has intensified regarding the human rights risks associated with their non-medical applications. In particular, scholars have highlighted privacy threats arising from the collection and use of neural data,[10] including risks to freedom of thought[11] and freedom of expression,[12] while stressing the

---

[2] D. Lupton, *The Quantified Self*, Cambridge, 2016, 2-3.

[3] UNESCO, *Final Draft Recommendation on the Ethics of Neurotechnology* (hereinafter *UNESCO Recommendation*), Paris, 2025, 4.

[4] L. Bernaez, V. Mahieu, *Neurotech consumer market atlas. How the sector is making moves into the mainstream,* in *Center for Future Generations*, 2025. Available at: https://cfg.eu/neurotech-market-atlas/#subchapter-6.

[5] Through electroencephalography (EGG) (Electroencephalogram), in *Mayo Clinic*, or functional magnetic resonance (fMRI). FMRi is a non-invasive method for studying the functional anatomy of the human brain. International Bioethics Committee (IBC) of UNESCO, *Report of the International Bioethics Committee of UNESCO (IBC) on the Ethical Issues of Neurotechnology* (hereinafter *IBC Report*), Paris, 2021, 8.

[6] M. Sosa Navarro, A. Lavazza, M. Balconi, M. Ienca, F. Minerva, F. Pizzetti, M. Reichlin, F. Samorè, V.A. Sironi, S. Songhorian, *Neuralink's brain-computer interfaces: medical innovations and ethical challenges*, in *Frontiers in Human Dynamics*, 7, 2025.

[7] Both through transcranial direct stimulation (TDCs) or transcranial magnetic stimulation (TMS). Transcrannial direct current stimulation (tDCS) or Transcranial electrical stimulation (tES) involve devices delivering continuous currents supposedly to enhance concentration or relaxation. *IBC Report*, cit. Transcranial magnetic is a non-invasive tool for the electrical stimulation of neural tissue, including cerebral cortex, spinal roots, and cranial and peripheral nerves. M. Kobayashi, *Transcrannial magnetic stimulation in neurology*, in *The Lancet Neurology*, 2(3), 145-156.

[8] C. Bublitz, S. Ligthart, *The new regulation of non-medical neurotechnologies in the European Union: overview and reflection,* in *Journal of Law and the Biosciences*, 11(2), 2024, 14

[9] M. Sosa Navarro, S. Dura-Bernal, *Human Rights Systems of Protection From Neurotechnologies That Alter Brain Activity*, in *Drexel Law Review*, 15, 2023, 908-913.

[10] P. Kellmeyer, *Big Brain Data: On the Responsible Use of Brain Data from Clinical and Consumer-Directed Neurotechnological Devices,* in *Neuroethics*, 14, 2021, 87-90.

serious threats that neurotechnologies capable of altering neural activity may present to mental integrity and personal identity.[13] Echoing these concerns, the UN Human Rights Council has recently adopted a report from the Human Rights Advisory Committee which underscored neurotechnology's unique character and socially disruptive potential noting that such technologies generally "(a) enable the exposition of cognitive processes; (b) enable the direct alteration of a person's mental processes and thoughts; (c) bypass the individual's conscious control or awareness; (d) enable non-consensual external access to thoughts, emotions and mental states; (e) are nurtured by "neurodata", which are needed for their own functioning, calibration and optimization; and (f) collect, analyse and process large personal datasets of a highly sensitive nature".[14] The heightened awareness expressed by this UN report, coupled with a global neurotechnology market projected to hit $24.2 billion by 2027,[15] has triggered the adoption of numerous policy reports and recommendations at regional and international levels,[16] alongside a patchwork of domestic legislative initiatives.[17]

Against this background, the deployment of neurotechnologies to augment workplace productivity by measuring and enhancing workers' concentration levels and cognitive performance poses

[11] C. BUBLITZ, *Freedom of Thought in the Age of Neuroscience*, in *Archiv für Rechts und Sozialphilosophie,* 100(1), 2014.

[12] S. LIGTHART, *Freedom of thought in Europe: do advances in "brain-reading" technology call for revision?*, in *Journal of Law and the Biosciences*, 7(1), 2020.

[13] S. LIGTHART, M. IENCA, G. MEYNEN *et al.*, *Minding Rights: Mapping Ethical and Legal Foundations of Neurorights,* in *Cambridge Quarterly of Healthcare Ethics,* 32(4), 2023.

[14] UNGA, Human Rights Council, *Impact, opportunities and challenges of neurotechnology with regard to the promotion and protection of all human rights*, A/HRC/57/61, 8 August 2024, 2-3. This report follows from a HRC resolution from 2022 that specifically requests the Advisory Committee to examine the risks and challenges arising from neurotechnologies. UN, Human Rights Council, *Res. 51/3. Neurotechnology and human rights*, Doc. No. A/HRC/RES/51/2, 13 October 2022.

[15] M. SQUICCIARINI, L. XU, *Unveiling the Neurotechnology Landscape: Scientific Advancements, Innovations and Major Trends*, UNESCO, Paris, 2023, 9. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000386137. Last visited 25/07/2025.

[16] See, among others, OECD, *Recommendation of the Council on Responsible Innovation in Neurotechnology*, OECD/LEGAL/0457, 11 December 2019; *UNESCO Recommendation,* cit.; Human Rights Council Advisory Committee, *Impact, Opportunities and Challenges of Neurotechnology with regard to the Promotion and Protection of All Human Rights*. Report of the Advisory Committee, UN Human Rights Council/UN General Assembly, 8 August 2024. For a full analysis of soft law instruments that have emerged in the last decade see chapter 6 of M. SOSA NAVARRO, *The role of soft law in the regulation and governance of human rights challenges posed by neurotechnologies*, Torino, 2025, 140 ff.

[17] In this sense, while Chile was the first country to amend its Constitution to protect psychological integrity and brain activity. Others have followed in Latin America with discussions underway in both Mexico and Brazil. D. BORBÓN, *Challenges of the inconsistent neurorights framework in Latin America,* in *Nat Neurosci*, 28, 2025, 1363-1364. Such a trend can also be identified in the Global North. Notably, France — through its Bioethics Law of 2 August 2021 — added Article L.1151-4 to the Public Health Code, prohibiting "any acts, procedures, techniques, methods or equipment that modify brain activity and pose a serious or suspected serious risk to human health". In light of that enactment, Article 16-14 of the French Civil Code was revised to provide that "brain-imaging techniques may only be used for medical or scientific research purposes, or in the context of a judicial expert appraisal, expressly excluding, in that context, functional brain imaging (fMRI)". In the US, States like Colorado (General Assembly 2024), California (SB 1223, 2024) or Montana (Senate Bill 2025) have adopted binding regulation to protect the privacy of neural data in response to the rapid development and market proliferation of neurotechnology. See M. SOSA NAVARRO, *Los neuroderechos en el norte global* in *Temas Selectos de Neuroética*, Ed. Aranzadi, Pamplona (*forthcoming*).

unprecedented challenges for the international human-rights framework.[18] Assessing their current and potential use in employment settings therefore demands not only a review of overarching human-rights instruments but also a close examination of sector-specific regimes, including international data-protection treaties and relevant documents adopted within the International Labour Organisation (ILO). This article undertakes a transversal analysis of international law's capacity to safeguard against infringements of human and labour rights arising from market-driven cognitive and emotional monitoring and enhancement.[19]

The paper is structured in three substantive parts. Section 2 examines how the rights to privacy and freedom of thought provide safeguards against workplace neurotechnologies, with particular focus on neural data protection and the interpretive developments of this notion within international law. Section 3 turns to the right to mental integrity, addressing the distinctive challenges posed by brain-altering neurotechnologies and the potential role of international labour standards. Section 4 explores the European regulatory framework, analysing the GDPR, the Medical Devices Regulation, and the AI Act to assess how regional instruments confront the risks of neuromonitoring and neuromodulation in employment contexts.

2. International human rights systems of protection against workplace surveillance

The recent proliferation of the so-called "bossware", which comprehends productivity monitoring systems such as software that tracks workers' keystrokes, breaks, and screen activity, marks the onset of new era of workplace surveillance with serious human-rights implications.[20] In this same context, wearable neurotechnologies have moved beyond niche safety workplace applications (such as monitoring fatigue levels in lorry drivers)[21] into everyday use, particularly within the consumer wellness market. Miniaturized EEG sensors embedded in headbands or earbuds, powered by machine-learning-based AI systems, are now marketed for purposes such as enhancing concentration, reducing stress, and monitoring attention levels.[22] These devices do not only monitor our cognitive and emotional states but can also, via closed-loop feedback, alter them. If used without stringent safeguards in environments characterised by power imbalances and a quest for higher productivity, such as the workplace, these

[18] S. ALEGRE, *Rethinking freedom of thought for the 21st Century,* in *European Human Rights Law*, 3, 2017, 232; J.M. MUÑOZ, J.A. MARINARO, *Neurorights as reconceptualized human rights*, in *Frontiers in Political Science*, 2023, 2.

[19] J.M. PEAKE, G. KERR, J.P. SULLIVAN, *Critical Review of Consumer Wearables, Mobile Applications, and Equipment for Providing Biofeedback, Monitoring Stress, and Sleep in Physically Active Populations,* in *Frontiers in Physiology*, 9, 2018, 743.

[20] A. ALOISI, V. DE STEFANO, *Essential jobs, remote work and digital surveillance: Addressing the COVID-19 pandemic panopticon,* in *International Labour Review,* 161, 2022, 289. Such a concern is illustrated in Ken Loach's recent film "Sorry we missed you" and has led to the so-called Amazonian Era, a name inspired by the technology company Amazon, that often spearheads illegitimate collection of worker's personal information. A. GILBERT, A. THOMAS, *The Amazonian era. How algorithmic systems are eroding good work,* 2021. See also S. ZUBOFF, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power,* New York, 2019.

[21] N. FARAHANY, *The Battle for your Brain*, New York, 2023, 41; J. LAROCCO, M. DONG LE, D.G. PAENG, *A Systemic Review of Available Low-Cost EEG Headsets Used for Drowsiness Detection*, in *Frontiers in Neuroinformatics*, 14, 2020.

[22] L. BÉRNAEZ TIMÓN, V. MAHIEU, *Neurotech consumer market atlas. How the sector is making moves into the mainstream*, in *Centre for future generations*, 2025, 18-20.

technologies may pose serious threats to employees' dignity by undermining privacy, freedom of thought, and mental integrity, while also affecting their sense of identity and personal agency.[23]
Building on Istace's distinction between neurotechnologies that merely collect neural data and those that actively reshape neural activity, this section proceeds in two complementary strands. First, it assesses whether existing safeguards for privacy at an international level can protect employees against the novel risks posed by neurotech-driven workplace monitoring.[24] Second, it explores whether the right to freedom of thought can serve as a brake on market-driven workplace surveillance.

2.1. Privacy as a precondition: special focus on Convention 108+ of the Council of Europe and Article 17 ICCPR

The International Law Commission (ILC) reminds us that privacy and personal data protection is all but a new concern to the international community.[25] It is thus unsurprising that, as Della Morte observes, this field benefits from a highly developed system of international and regional protections.[26] International human-rights law long recognized privacy as an essential facet of the broader right to private life, enshrined in Article 12 of the UDHR, Article 17 of the ICCPR, Article 8 of the ECHR and Article 11 of the ACHR. Notably, in its General Comment 16 on Article 17 ICCPR, the UN Human Rights Committee affirmed that protections against "arbitrary or unlawful interference" with privacy extend not only to State action but also to conduct by non-State actors, including private entities.[27] This broadened interpretation is especially pertinent in light of escalating workplace surveillance, now potentially encompassing neurotechnologies and "bossware" capable of collecting and processing employees' neural data and making inferences on their cognitive and emotional states from it.[28] Such developments have prompted scholars like Malgieri and Ienca to argue that the unique sensitivity of neural data demands safeguards tailored to the specific risks of cognitive surveillance, manipulation, and discrimination.[29]

---

[23] P. KELLMEYER, *Big Brain Data,* cit*.,* 1

[24] T. ISTACE, *Human rights law: an incomplete but flexible framework to protect the human mind against neurotechnological intrusions*, in *Law, Innovation and Technology*, 16(1), 2024, 18.

[25] It has actually been a concern "since the late 60s". International Law Commission, *Report of the International Law Commission on the work of its Fifty-eighth session, 7 May-8 June and 9 July-10 August 2006*, UN GAOR, 61st sess, Supp No 10, UN Doc A/61/10. *D. Protection of Personal Data in Transborder Flow of information,* 493.

[26] G. DELLA MORTE, *Big Data e Protezione Internazionale dei Diritti Umani*, Naples, 2018, 271-273. For a comprehensive analysis of these protective frameworks — which is further enriched by extensive ECtHR and CJEU case law and a suite of soft-law principles summarized by the ILC in eleven overarching tenets — see Chapter I of Part II in the authoritative volume cited above.

[27] UN Human Rights Committee, *General Comment No. 16: Article 17 (Right to Privacy), The Right to Respect of Privacy, Family, Home and Correspondence, and Protection of Honour and Reputation*, 8 April 1988, UN Doc. HRI/GEN/1/Rev.9 (Vol. I), para. 1

[28] Neurotechnology based tools sold as "EEG tech to aid workplace wellness" are already available in the market. See for instance, Emotiv or SmartCaps, an EEG-based technology currently used as safety equipment for fatigue monitoring.

[29] G. MALGIERI, M. IENCA, *Mental data protection and the GDPR*, in *Journal of Law and the Biosciences*, 9(1), 2022, 2, 8 and 10. In the same line, other scholars had advocated for a human rights impact assessment for AI, a proposal that was ultimately introduced in the AI Act for high-risk systems. A. MANTELERO, *Human Rights Impact Assessment and AI* in *Beyond Data*, in *Information Technology and Law Series,* 36, 2022.

Against this backdrop, Convention 108+, the modernized successor to the 1981 Council of Europe Data Protection Convention, offers a particularly promising framework. As the first binding international treaty devoted to personal data protection, its amended form, scheduled to enter into force sometime in 2026,[30] interestingly extends its scope to data processed for national security and defence,[31] an innovation of direct relevance to dual-use or converging technologies such as the one under consideration.[32] The choice is especially noteworthy given that legislative initiatives regulating AI do not necessarily extend to military applications, as illustrated by the EU AI Act, an omission that some commentators have described as one of its "most glaring oversights".[33]

Under Convention 108+, the collection and processing of personal data must satisfy seven core principles: transparency; purpose specificity; necessity and proportionality; data minimization and accuracy; secure processing; and respect for data-subject rights of access, rectification, objection and erasure. However, these obligations apply only to data that remain 'personal' in the sense of enabling identification or re-identification. Followingly, neural information that has been fully anonymized currently falls outside this regime and would require bespoke protective measures.[34] It is, however, important to note that although the Convention establishes a qualified category of personal data, emotions and thoughts not linked to health status, sexuality, or political and religious beliefs are not explicitly included in this category and are thus not granted specific protection.[35] This oversight has not gone unnoticed within the international scholarly community. At the 46th Plenary Meeting of the Council of Europe, held on 5 June 2024, international legal scholar Edoardo Bertoni and bioethicist Marcello Ienca presented a paper which, after reviewing the biological, legal, and ethical foundations of the risks to mental privacy posed by the proliferation of non-medical neurotechnologies, urged Member States to adopt a risk-based classification system. Central to their proposal was the introduction of a *"Mental*

[30] At the time of writing only 33 out of the 38 required States for the Convention to enter into force have ratified Convention 108+. See updated chart of signatures and ratifications here: https://www.coe.int/en/web/conventions/full–list?module=signatures–by–treaty&treatynum=223 (Last accessed 26/01/2026).

[31] Previously, Article 3(2)(a) permitted exemptions for data collection and processing carried out in the national security and intelligence domain. By contrast, the GDPR expressly excludes any processing for national security or intelligence purposes and the proposed AI Act likewise omits coverage of these activities. See R. JANSEN, M. REIJNEVELD, *Convention 108+, the GDPR, and Data Processing in the National Security Domain*, in *EUR. DATA PROT. L. REV*, 8, 2022, 423.

[32] A comprehensive examination of the challenges that converging technologies such as neurotechnologies pose to trade regulation, both at the international and regional levels, lies beyond the scope of this paper. For an in-depth discussion see Chapter 5 (Dual use neurotechnology in the era of technological convergence) of M. SOSA NAVARRO, *The role of soft law in the regulation and governance of human rights challenges posed by neurotechnologies,* cit. See also J.M. RICKLI, M. IENCA, *The Security and Military Implications of Neurotechnology and Artificial Intelligence,* in O. FRIEDRICH, A. WOLKENSTEIN, C. BUBLITZ, R.J. JOX, E. RACINE, (eds), *Clinical Neurotechnology meets Artificial Intelligence. Advances in Neuroethics*, Cham, 2021.

[33] O. POLLICINO, F. PAOLUCCI, *Regulating AI Autonomy: A Constitutional Framework for the Digital Era,* in M. DURANTE, U. PAGALLO (eds.), *Handbook on Law and Digital Technologies*, de Gruyter, forthcoming, Bocconi Legal Studies Research Paper No. 5098433, Milan, 2024, 16.

[34] For a full-fledged distinction between anonymous and pseudoanonymous data, pursuant to the innovative framework introduced by the GDPR see G. DELLA MORTE, *Big Data e Protezione Internazionale,* cit., 156-157.

[35] *Ivi,* 26. This may fail to grant protection to data collected about these internal processes when they are not related to the cited categories, that is, when they broadly identify thoughts and feelings.

Data Protection Impact Assessment", modelled on the AI Act's fundamental-rights impact assessment.[36] Such an approach would constitute a first step towards aligning international data-protection law with the UN Guiding Principles on Business and Human Rights, embedding substantive human-rights due diligence into the core of neurotechnology governance.

To date, both Convention 108 and its modernised version, Convention 108+, have prompted domestic legislative reforms across Europe to bring national laws into closer alignment with their principles. Certain safeguards and standards contained in these instruments, specifically the requirements of proportionality and necessity, have informed the reasoning of two landmark judgments of the European Court of Human Rights.[37] Nonetheless, the Conventions have not shaped the Court's privacy jurisprudence to the same extent as the Oviedo Convention on Human Rights and Biomedicine has. This instrument, also adopted under the Council of Europe, has significantly influenced the ECtHR's case-law with regards to the ethical challenges arising from advances in genetics, particularly genome sequencing and editing technologies.[38] Within the specific context of employment, prior to the modernization of Convention 108, the Council of Europe had already addressed the protection of personal data in the workplace through its Recommendation of 18 January 1989, subsequently updated by Recommendation CM/Rec (2015)5 of 1 April 2015. Notably, this instrument stresses the desirability of extending to the employment sector the principles enshrined in Convention 108 and its Additional Protocol concerning supervisory authorities and transborder data flows. At the same time, it explicitly underscores that the processing of personal data must respect human dignity and privacy so as to 'allow for the free development of the employee's personality as well as for possibilities of individual and social relationships in the workplace' (para. 3). While explicitly prohibiting the use of information systems and technologies whose direct and principal purpose is to monitor employees' activity and behaviour (para. 15), the Recommendation nevertheless acknowledges that certain monitoring-adjacent technologies may be introduced for legitimate organisational objectives. In such cases, their deployment must be preceded by consultation with employees' representatives and comply with the additional safeguards set out in Principle 21, reflecting an awareness of the potentially significant indirect effects of workplace monitoring on workers' rights.[39]

---

[36] *Ivi*, 30.

[37] ECtHR*, Centrum för Rättvisa v. Sweden*, Application no. 35252/08, Judgment (Grand Chamber), 25 May 2021, Strasbourg; ECtHR*, Big Brother Watch and Others v. the United Kingdom*, Applications nos. 58170/13, 62322/14 and 24960/15, Judgment (Grand Chamber), 25 May 2021, Strasbourg.

[38] F. SEATZU, *The Experience of the European Court of Human Rights with the European Convention on Human Rights and Biomedicine* in *Utrecht Journal of International and European Law*, 31(81), 2015. For a specific analysis of the Oviedo Convention's influence on the Strasbourg Court's interpretation of the right to family life and the content of the right to informed consent see 10-11. G. CATALDI, *La Convenzione sui Diritti Umani e la Biomedicina*, in L. PINESCHI (ed.), *Tutela Internazionale dei Diritti Umani*, Giuffrè, Milano, 2006, 589. For a comprehensive examination of this instrument see also R. SAPIENZA, *La Convenzione europea sui diritti dell'uomo e la biomedicina,* in *Rivista di Diritto Internazionale*, 1998, 457-470.

[39] For a detailed analysis of the content of this Recommendation see A. SARTORI, *Il controllo tecnologico dei lavoratori, La nuova disciplina italiana tra vincoli sovranazionali e modelli comparati,* Milan, 2020, 8-9. The additional safeguards set out in Principle 21, and whose respect should be ensured by the employers include: a. informing employees before the introduction of information systems and technologies enabling the monitoring of their activities. (…); b. taking appropriate internal measures relating to the processing of that data and notify employees in advance; c. consult employees' representatives in accordance with domestic law or practice, before

It can be inferred from the above that despite the significant efforts undertaken by the Council of Europe, the limited protection of thoughts and emotions, combined with the Convention's regional scope, highlight important shortcomings in the only binding international instrument on personal data protection currently in force. This gap reinforces the role of Article 17 of the ICCPR as the principal normative anchor at the global level and underscores the urgent need for a coherent and consensual interpretation of the right to privacy that can encompass the emerging dimension of mental privacy. Such an effort was undertaken in part by the UN Special Rapporteur on the Right to Privacy, Ana Brian Nougrères, who between 2022 and 2025 issued a series of reports, two of which are of particular relevance for the purposes of this paper.[40]

The first report,[41] adopted in 2022, articulates the normative foundations of privacy and personal data protection, identifying ten guiding principles[42] as the core components of what the Special Rapporteur characterizes as a global regulatory architecture. According to Nougrères, these principles serve a dual function: they both guide the interpretation of the existing normative framework and facilitate its consistent application across diverse contexts. The Rapporteur goes a step further when asserting that these principles are not to be treated as mere recommendations, but as essential benchmarks for the lawful and ethical processing of personal data.[43] The principle of purpose specification requires that data be collected solely for explicit and legitimate purposes, with repurposing prohibited if incompatible with the original aim. The principle of transparency demands that the objectives of processing, along with the identity and contact details of controllers or their representatives, be disclosed at all times.[44] Consent, widely recognized at the international level as a legal ground for processing and closely tied to the principle of legality, must be freely given, specific, informed, unambiguous, and revocable.[45] Under the principles of data minimization and proportionality, only data strictly necessary for the stated purpose may be processed, and such processing must be proportionate to the aim pursued. Proportionality,

---

any monitoring system can be introduced or in circumstances where such monitoring may change. (…); d. consult, in accordance with domestic law, the national supervisory authority on the processing of personal data.

[40] It is worth noting that, although neither of these two documents made explicit reference to workers, the right to privacy had already been invoked at the UN level in General Assembly Resolution 45/95 of 14 December 1990, which adopted the *Guidelines for the Regulation of Computerized Personal Data*, and more recently in Resolution 68/167 of 18 December 2013 on the right to privacy in the digital age. Although this marked the first step in placing privacy on the UN digital agenda, led by the General Assembly, significant developments have followed. These include a series of reports by the Office of the High Commissioner for Human Rights (OHCHR, 2014, 2018, 2022) addressing surveillance, big data, AI, and digital identity, as well as the establishment in 2015 of a dedicated Special Rapporteur on the right to privacy, whose successive mandates have produced extensive thematic reports covering, inter alia, state and corporate surveillance, algorithmic management, and most recently neurotechnologies, several of which will be next examined.

[41] A.B. NOUGRÈRES, *Principles Underpinning Privacy and the Protection of Personal Data, Report to the UN Human Rights Council*, UN Doc A/77/196 (20 July 2022).

[42] These principles are legality, lawfulness and legitimacy, consent, transparency, purpose specification, fairness, proportionality, data minimization, data quality, accountability, and security.

[43] A.B. NOUGRÈRES, *Principles Underpinning Privacy and the Protection of Personal Data, Report to the UN Human Rights Council*, UN Doc A/77/196 (20 July 2022), cit., 3.

[44] A.B. NOUGRÈRES, *Principles Underpinning Privacy and the Protection of Personal Data, Report to the UN Human Rights Council*, UN Doc A/77/196 (20 July 2022), cit., 11.

[45] A.B. NOUGRÈRES, *Principles Underpinning Privacy and the Protection of Personal Data, Report to the UN Human Rights Council*, UN Doc A/77/196 (20 July 2022), cit., 8.

moreover, requires the controller to use the processing operation that is least invasive in terms of privacy.[46]

Yet, the recognition of these principles has not removed the qualified nature of the right to privacy under international and regional human rights law. In practice, privacy is generally conceived as a right subject to limitations, provided that restrictions are lawful, necessary, and proportionate.[47] This has particular relevance in safety-critical sectors (e.g., transport, aviation, healthcare), where interferences with employee privacy may be justified to prevent serious harm,[48] such as through sleep reporting, alertness monitoring, or real-time drowsiness detection for drivers.[49]

Despite not enjoying the protections granted to absolute rights, the Special Rapporteur has underscored that neural data present distinctive challenges that go beyond traditional privacy considerations. In her most recent report, published in January 2025, Nougrères examines the foundations and principles for regulating neurotechnologies and the processing of neural data through the lens of the right to privacy.[50] She highlights that the heightened sensitivity of neural data derives not only from its capacity to identify individuals but also from its ability to reveal cognitive and affective states and to reflect personal experiences and emotions.[51] In light of these features, the Special Rapporteur proposes the creation of a Model Law that serves as a tool to harmonize domestic regulations by establishing minimum standards for the safe and ethical use of neurotechnologies.[52] Such Model Law, should not only integrate the general principles governing privacy and data protection in general but should also include a set of neuro-specific requirements.[53] In addition to demanding that any processing be necessary and narrowly tailored, the 2025 report calls for: (i) a human-rights and human-dignity-based approach across the design, development, deployment, commercialization and use of neurotechnologies; (ii) recognition of neurodata as highly sensitive given its capacity to reveal cognitive and affective states; (iii) application of the precautionary principle and safety-by-design; and (iv) rights-protective governance instruments (including stricter transparency, oversight and effective remedies). Taken together, these standards establish a heightened threshold for the regulation of neurotechnology-based monitoring in the workplace. A human-dignity based approach would thus call for the revision of the role attributed to employees' consent to provide access to his/her neural data in the workplace. As some authors have noted, the inherently imbalanced nature of the employment relationship renders employee consent an insufficient safeguard, thereby necessitating additional

[46] A.B. Nougrères, *Principles Underpinning Privacy and the Protection of Personal Data, Report to the UN Human Rights Council*, UN Doc A/77/196 (20 July 2022), cit., 15.

[47] ECtHR, *Segerstedt-Wiberg and Others v. Sweden*, Application no. 62332/00, Judgment (Chamber), 6 June 2006, para 88.

[48] N. Farahany, *The Battle for your Brain*, cit., 41.

[49] P.M. Ramos, C.B. Maior, M.C., Moura, I. D. Lins, *Automatic drowsiness detection for safety-critical operations using ensemble models and EEG signals,* in *Process Saf. Environ. Prot.*, 164, 2022, 566-581.

[50] A.B. Nougrères, *Foundations and Principles for the regulation of neurotechnologies and the processing of neurodata from the perspective of the right to privacy, Report to the UN Human Rights Council,* UN Doc A/HRC758/58 (15 January 2025).

[51] *Ivi*, 6.

[52] *Ivi*, 7.

[53] *Ivi*, 9-10.

protective measures.[54] Failure to comply with these requirements risks not only infringing privacy but also eroding workers' autonomy and dignity.

In the absence of a specific regulatory framework governing the workplace use of EEG-enabled earbuds, headbands, or smart glasses, the growing deployment of digital monitoring tools that collect, process, and interpret neural data exposes workers' right to mental privacy to significant risks. By enabling employers to monitor attention, emotional states, and cognitive patterns, these technologies leave mental privacy particularly vulnerable. This vulnerability creates a concrete risk of discrimination, as neurotechnologies may enable employers to make hiring, promotion, or incentive decisions based on performance-related, emotional, or health-related information inferred from neural data.

Ultimately, while Article 17 of the ICCPR provides a foundational safeguard against arbitrary or unlawful interferences with privacy, including by non-State actors, the unique characteristics of neural data call for an updated interpretation of its scope to address the specific and evolving risks of neurotechnology-based surveillance, particularly in the workplace. As emphasised by the UN Special Rapporteur on the right to privacy in her report on neural data, these risks highlight the need to reconceptualise the protection of mental privacy. Although the Special Rapporteur's reports are formally non-binding, they extend beyond interpretive guidance: they provide evidence of an emerging global practice that may contribute to the formation of customary standards in data protection and privacy by both interpreting the existing framework and harmonising its application across diverse regulatory contexts. As such, Article 17 ICCPR calls for further normative development and interpretive expansion, particularly through the Human Rights Committee's jurisprudence, and most notably its General Comments, to provide a framework capable of regulating the emerging "digital mind" paradigm in workplace settings and ensuring that workers' mental privacy is not violated under the guise of productivity or safety.

2.2. The right to freedom of thought in the workplace under international human rights law

While the preceding section framed respect for privacy as a precondition for safeguarding workers' rights in environments increasingly permeated by monitoring practices, this section turns to the right to freedom of thought. The extent to which this right may provide protection against violations stemming from the use of neurotechnological monitoring tools for efficiency, performance or well-being purposes is examined.

The right to freedom of thought (RFoT) enjoys longstanding protection under core international and regional human rights instruments (Article 18 of the UDHR and the ICCPR, Article 9 of the ECHR, and Article 13 of the ACHR). Traditionally, this right has been interpreted as safeguarding the external manifestations of thought, such as speech or conduct, rather than the internal cognitive processes themselves. However, the proliferation of neurotechnologies capable of accessing, inferring, or even altering mental states has prompted renewed scholarly and institutional efforts to reconceptualize the RFoT in light of these developments.[55]

[54] H. ABRAHA, *A pragmatic compromise? The role of Article 88 GDPR in upholding privacy in the workplace,* in *International Data Privacy Law*, 12 (4), 2022, 294.

[55] C. BUBLITZ, *Freedom of Thought in the Age of Neuroscience*, cit.; S. MCCARTHY-JONES, *The Autonomous Mind: The Right to Freedom of Thought in the Twenty-First Century,* in *Front. Artif. Intell.,* 25(1), 2019; S. ALEGRE, *Rethinking*

Two main approaches have emerged. The first calls for an explicit expansion of the RFoT to include protection of the internal dimension of thought, what some refer to as cognitive liberty.[56] The second proposes adding a relative component to this traditionally absolute right in order to strengthen it and stimulate its application in practice.[57] From this perspective, the development of narrowly tailored implied limitations to certain aspects of the right, such as the freedom against impermissible alteration of thought, may be legally grounded in the understanding that inner thoughts are not entirely beyond the reach of state power, whether from a factual, epistemic, or normative standpoint.[58] This reconceptualization acquires particular relevance in the context of neurotechnologies and carries significant implications for their use in workplace settings, where access to inner thoughts may extend to both public and private employers.

As the UN Special Rapporteur on Freedom of Religion or Belief, Ahmed Shaheed, outlined in his 2021 report, the RFoT should be understood as multidimensional. These dimensions include: (1) the right not to disclose one's thoughts; (2) freedom from punishment based solely on one's thoughts; (3) protection from impermissible interference or alteration of thought; and (4) the creation of an enabling environment for free thinking.[59] Each of these dimensions is potentially jeopardized by the introduction of cognitive-monitoring tools in professional settings.[60] For instance, when employers deploy technologies capable of decoding levels of attention, emotional states, or mental fatigue, they may indirectly compel disclosure of internal mental states, violating the first dimension. Even in the absence of explicit coercion, the existence of such surveillance tools may foster a chilling effect where workers self-censor their thoughts or emotions in anticipation of scrutiny (second dimension).[61]

In line with the aforementioned discussion, experts from international organizations and academia advocate for using the precautionary principle to guide the regulation of neurotechnologies.[62] This principle, long recognized in international environmental and health law as having crystallized into customary law,[63] is particularly suited to contexts where scientific uncertainty intersects with high-stakes risks for individual rights and societal well-being.[64] It allows regulators to take protective action in the face of credible threats, even in the absence of conclusive scientific proof.

---

freedom of thought for the 21st Century, cit.; S. LIGTHART C. BUBLITZ, T. DOUGLAS, L. FORSBERG, G. MEYNEN, *Rethinking the Right to Freedom of Thought: A Multidisciplinary Analysis,* in *Hum. Rts. L. Rev. 1*, 3, 2022.

[56] M. IENCA, *On neurorights,* in *Frontiers in Human Neuroscience*, 15, 2021; N. FARAHANY, *The Battle for your Brain*, New York, 2023.

[57] S. LIGTHART, *Reconsidering the absolute nature of the right to freedom of thought,* in *Human Rights Law Review*, 25 (3), 2025, 26.

[58] *Ivi*, 5-6.

[59] A. SHAHEED (Special Rapporteur on Freedom of Religion or Belief), *Interim Report on the Freedom of Thought*, 14, U.N. Doc. A/76/380, Oct. 5, 2021, para. 25.

[60] A. MCCAY, *Neurotechnology, law and the legal profession*, cit., 5.

[61] K. BALL, *Electronic Monitoring and Surveillance in the Workplace. Literature review and policy recommendations', Publications Office of the European Union, 2021,* 34-40 for the individual level of analysis of the effect of the impact of surveillance/monitoring systems in the workplace. See also 71 and 78.

[62] *See* OECD, *Recommendation of the Council on Responsible Innovation in Neurotechnology,* cit., 6-9.

[63] C. RAGNI, *Scienza, Diritto e Giustizia Internazionale,* Naples, 2020, 103 ff.

[64] N. DE SADELEER, *Environmental principles. From political slogans to legal rules,* Oxford, 2002, *Chapter 3. The precautionary principle*, in particular, 174 ff.

From another perspective, the use of such technologies undermines the enabling environment essential for the autonomous development of thought. Drawing from Shaheed's framing, the freedom to think does not merely require the absence of interference, it requires the certainty that one's mental processes remain fully private during the critical "rumination phase" of thought formation (fourth dimension).[65] This concern is magnified in employment settings marked by power asymmetries,[66] where the unregulated use of neurotechnologies may undermine workers' effective ability to exercise their rights, particularly if neural data is used for productivity metrics or behavioral profiling. While national constitutional protections may offer pathways for enforcement and remedy in case of violation, international law remains underdeveloped in this area. The current absence of clear regulatory boundaries means that the normative dimensions of RFoT, especially in the workplace, remain inadequately protected.

Ultimately, this section has argued that protecting workers' mental privacy and freedom of thought in the age of neurotechnology demands both innovative legal reasoning and the dynamic reinterpretation of existing rights frameworks. In particular, it underscores the pressing need for robust safeguards against practices that compel, incentivize, or normalize the disclosure of internal mental states for commercial or productivity-related purposes. Crucially, the positive obligations arising from so-called "negative" rights, consistently recognized in human rights jurisprudence,[67] must be fully acknowledged in this context. These obligations require not only refraining from unlawful interference,[68] but also adopting proactive measures, such as enacting regulatory frameworks, to prevent non-state actors from infringing upon individuals' rights.[69] Some of these concerns, particularly those relating to surveillance and working conditions, will be examined in the following section, which considers the protections afforded to workers under International Labour Organization (ILO) conventions against human rights violations arising from surveillance practices.

3. The International Labour Organization and emerging challenges of workplace surveillance

3.1. ILO safeguards of mental privacy

The International Labour Organization (ILO) is among the oldest existing international organisations, tracing its origins to 1919 when it was established as part of the Treaty of Versailles to "reflect the belief

[65] N. FARAHANY, *The Costs of Changing Our Minds*, in *Emory L.J.,* 69, 2019, 98.

[66] The European Court of Human Rights has emphasized that the validity of consent hinges on the existence of a "real choice". It has held that even when individuals disclose personal data voluntarily or with consent, such disclosure does not strip them of the protections under Article 8 ECHR if no real choice exists – such as when an employer requires the disclosure of a job-seeker's criminal record as a condition of employment. ECtHR, *M.M v. United Kingdom*, Application no. 24029/07, Judgment of 13 November 2012, para. 189.

[67] V. STOYANOVA, *The Disjunctive Structure of Positive Rights under the European Convention on Human Rights,* in *Nordic Journal of International Law,* 87(3), 2018, 345. See specifically footnote 1.

[68] S. LIGTHART, *Freedom of thought in Europe: do advances in 'brain-reading' technology call for revision?*, cit., 18 and 15-16 noting the range of case-law interpreting Article 9 of the ECHR, that only recent cases "provide a bit more clarification", and that, in general, "case-law and decisions do not extensively elaborate on the meaning and scope of the notion of thought as protected by Article 9 ECHR".

[69] S. ALEGRE, *Rethinking freedom of thought for the 21st Century*, cit., 222.

that universal and lasting peace can be accomplished only if it is based on social justice".[70] International labour law is primarily composed of ILO Conventions and Recommendations, which differ in legal character: while Conventions create binding obligations for States that ratify them, Recommendations are non-binding and provide guidance for national policy, legislation, and practice. Both instruments are adopted at the International Labour Conference and must subsequently be submitted by all 187 Member States to the competent domestic authority for consideration. Ratification of a Convention entails concrete effects at the national level, as enterprises become directly subject to laws, regulations, judicial decisions, and collective agreements implementing international labour standards.

To date, the only instrument specifically addressing the protection of workers' privacy within the ILO framework is the *Code of Practice on the Protection of Workers' Personal Data*, adopted at a meeting of experts on privacy held from 1-7 October 1996.[71] Unlike Conventions or Recommendations adopted under Article 19 of the ILO Constitution, Codes of Practice are non-binding technical standards developed through expert meetings convened by the ILO Governing Body. Drafted within the Organisation's tripartite structure, they involve representatives of governments, employers and workers. Once adopted, they are published by the ILO as practical guidance intended to assist in the implementation of existing standards. Their status is therefore that of soft law: they do not constitute formal sources of international labour law, nor are they subject to supervision by the ILO's Committee of Experts on the Application of Conventions and Recommendations (CEACR). Nevertheless, CEACR reports occasionally cite Codes of Practice as interpretive aids, and in practice such instruments often serve as important reference points for the development of legislation, collective agreements, workplace regulations, and company practices,[72] illustrating the post-law function of soft law in guiding normative interpretation and development.[73]

Despite being almost 3 decades old, the ILO's *Code of Practice on the Protection of Workers' Personal Data* introduces interesting guidance with regard to technology-based monitoring of workers. While not prohibited, it requires that workers be informed in advance of the reasons, methods, timing, and scope of any monitoring, and that employers minimise intrusions into privacy. Secret monitoring is allowed only where authorised by national law or justified by reasonable suspicion of criminal activity or other serious misconduct, while continuous monitoring is restricted to circumstances necessary for health, safety, or property protection (para 6.14). The *Code of Practice* also addresses invasive testing practices (paras. 6.10-6.12), prohibiting the use of polygraphs or similar 'truth-verification' technologies, strictly limiting genetic screening to instances expressly authorised by law, and requiring that personality tests remain consistent with privacy protections while preserving the worker's right to object. These provisions, although formulated in the mid-1990s, resonate strongly with contemporary debates on neurotechnology in the workplace. Brain-reading devices, much like polygraphs or personality profiling, seek to access otherwise inaccessible aspects of a person's inner life (emotions, mental states, or

[70] G. CASALE, *Fundamentals of International Labor Law*, 4th Ed., Milan, 2024, 23.

[71] On this topic, S. GIUBBONI, *Potere datoriale di controllo e diritto alla* privacy *del lavoratore. Una sinossi delle fonti europee e internazionali*, in *Riv. giur. lav. prev. soc.*, 1, 2012, 81.

[72] F. HENDRICKX, *Employment privacy* in R. BLANPAIN (ed.), Co*mparative Labour Law and Industrial Relations* in *Industrialized Market Economies*, The Netherlands, 2014, 476.

[73] For an in-depth study of the triple function of soft law theory in international law see E. TRAMONTANA, *Il soft law e la resilienza internazionale,* in *Ars interpretandi*, 2, 2017.

predispositions) and therefore raise comparable, if not heightened, risks for privacy, dignity, and freedom of thought.

In this sense, the *Code of Practice* can be seen as an early attempt to grapple with the very concerns that neurotechnologies now amplify in workplace settings, suggesting that its framework could be updated and expanded to explicitly encompass the unique risks posed by contemporary brain-monitoring and cognitive surveillance tools.

3.2. ILO safeguards of mental integrity

The preceding sections have examined the protection of neural data and freedom of thought, showing that these aspects can be effectively anchored within the existing international human rights framework. However, when it comes to neurotechnologies capable of altering brain activity, a significant normative gap emerges. While medical applications of neurotechnology benefit from heightened legal and ethical scrutiny, the increasing use of wearable brain-altering devices for productivity enhancement purposes remains largely unregulated at a global level.[74] This asymmetry raises serious concerns regarding worker's safety, discrimination, mental health, and the preservation of mental integrity.

Given its primary focus on privacy and freedom of thought, this paper does not seek to offer a comprehensive analysis of the risks to mental integrity arising from the spread of wearable neurostimulation devices, nor does it provide an exhaustive account of the legal safeguards afforded to workers under international law in this area. It does, however, aim to clarify the practical distinctions between brain-reading and brain-altering neurotechnologies, examined here from conceptual, practical, and normative perspectives.

In their foundational articulation of neuro-rights, *Ienca and Andorno* define mental integrity as the right to be protected against unauthorized alterations of neural computation, particularly where such changes result in harm.[75] According to their framework, three conditions must be met for an action to constitute a threat to mental integrity: (i) direct access to and manipulation of neural signals, (ii) lack of informed consent from the individual, and (iii) resulting harm to mental or physical well-being, including, for instance the side-effects of neuromodulation, an increasing risk due to the growing number of wearable neurostimulators on the market.[76] Importantly, the requirement of an action (to alter, manipulate, erase) excludes neurotechnologies used solely for monitoring or inference, sufficiently covered by the right to privacy and the right to freedom of thought examined in the previous sections of this article.

Establishing the fulfilment of the second condition — the absence of informed consent — is particularly problematic in workplace settings, where structural power imbalances between employer and employee call into question the voluntariness of any purported consent. This dynamic heightens the risk of coercive or exploitative uses of consumer neurostimulation devices aimed at enhancing worker

[74] In section 4.2., we will see this is not the case at the EU level, which has recently amended the Medical Devices Directive to include non-invasive non-medical brain stimulation devices within in the highest risk category of medical devices.

[75] M. IENCA, R. ANDORNO, *Towards new human rights in the age of neuroscience and neurotechnology,* in *Life Sci Soc Policy*, 13(5), 2017, 18.

[76] A. WEXLER, P.B. REINER, *Oversight of Direct-to-Consumer Neurotechnologies,* in *Science,* 2019, 234.

performance. While minor physical effects like skin irritation may be addressed under general product safety regimes,[77] the more significant concern lies in their potential to modulate cognitive states without genuine, informed, and voluntary consent. This challenge becomes even more pressing given the growing accessibility of non-invasive neurostimulators, which now appear in forms as discreet as earbuds or headbands and are marketed as everyday productivity tools.[78]

The internationally recognized right to bodily integrity might provide some protection in this context. While some authors suggests that any interference with bodily or mental autonomy should be assessed by the severity of its impact,[79] others argue that any unconsented modification of neural functioning, even via external stimulation, constitutes a violation of bodily integrity.[80] Building on this reasoning, *Istace* emphasizes that non-invasive devices like transcranial direct stimulation (TDCs) or neurostimulation earbuds must still fall within the scope of the right to bodily and mental integrity. This author challenges the adequacy of traditional legal distinctions based on physical. This perspective acquires particular relevance in light of increasing empirical evidence of the measurable effects of these technologies on attention, memory, and mood.[81]

Against this background, unregulated neuromodulation in the workplace, particularly when deployed in contexts of implicit coercion, raises concerns not only for the right to privacy, but for the core rights to bodily and mental integrity and personal dignity.[82] Although the right to bodily integrity may play a central role in safeguarding mental autonomy against such interventions, its application within the employment context demands a more tailored analysis. The unique vulnerabilities of the workplace, including structural power imbalances already referred to and the pressure to enhance performance, call for a perspective that transcends the general international human rights framework.

A possible avenue to explore could be the normative acquis of the International Labour Organization (ILO). Existing instruments already contain principles that provide specific guardrails against exploitative or harmful workplace practices that could be extended to cover neurotechnologies Convention No. 155 (1981) on Occupational Safety and Health and Recommendation No. 164 impose duties on employers to protect workers' physical and mental well-being; Convention No. 187 (2006) encourages continuous national improvement in occupational safety regimes; Convention No. 190 (2019) addresses psychological harm and harassment, including that enabled by technology; and Convention No. 111 (1958) prohibits discrimination in employment, which is relevant to algorithmic or neuro-based evaluations of performance.

---

[77] A. WEXLER, *A Pragmatic Analysis of the Regulation of Consumer Transcranial Direct Current Stimulation (TDCS) Devices in the United States,* in *J.L. & BIOSCIENCES,* 2, 2015, 683. This is mostly the case in the US, In the EU, as section 4.2. will illustrate physical effects of non-medical applications of neurostimulation devices fall under the Medical Device Regulation.

[78] N. FARAHANY, *The Battle for your Brain*, cit., chap. 3.

[79] V. TESINK, T. DOUGLAS, L. FORSBERG *et al.*, *Neurointerventions in Criminal Justice: On the Scope of the Moral Right to Bodily Integrity*, in *Neuroethics*, 16, 2023, 26

[80] J. RYBERG, *Neurointerventions, Crime, and Punishment: Ethical Considerations* Oxford, 2019, 52.

[81] T. ISTACE, *Human rights law: an incomplete but flexible framework to protect the human mind*, cit., 18

[82] R. BLUHM, M. CORTRIGHT, ED ACHTYES, LY CABRERA, *They Are Invasive in Different Ways: Stakeholders' Perceptions of the Invasiveness of Psychiatric Electroceutical Interventions*, in *AJOB Neuroscience*, 14 (1), 2023.

In conclusion, the analysis underscores that while privacy and freedom of thought provide an important starting point for regulating workplace neurotechnologies, the advent of devices capable of altering neural activity exposes a profound normative lacuna in international law. The risks of coercion, discrimination, and compromised mental autonomy associated with consumer neurostimulation tools evidence that workers' mental integrity cannot be adequately safeguarded by existing privacy-oriented frameworks alone. Although the right to bodily integrity offers a potential avenue for protection, its application in the employment context remains insufficiently theorised and operationalised. The ILO acquis, with its established emphasis on health, safety, dignity, and non-discrimination in the workplace, could serve as a promising foundation for extending protection to this emerging frontier.

4. An overview of the human-centered but fragmented European normative framework protecting the worker's mental privacy and mental integrity

Following the preceding discussion on the capacity of international human rights law to safeguard workers' rights to privacy and freedom of thought in the light of emerging workplace surveillance trends, it is now useful to consider how the European regional framework addresses these concerns. This perspective provides a more concrete view of how such rights, enshrined in the European Convention on Human Rights, can be enforced through binding legal instruments. Within this context, particular attention will be given to the General Data Protection Regulation (GDPR), and especially Article 88, which permits Member States to introduce specific rules on data processing in the employment context, provided these rules incorporate suitable and specific safeguards for human dignity and fundamental rights.

In addition, reference will be made to the EU Medical Devices Regulation (MDR), which contains provisions for the marketing of neurostimulation devices, including those intended for non-medical purposes, such as cognitive enhancement or performance modulation, that may be used in the workplace. These safeguards provide a legal foundation for regulating potentially harmful neurotechnological interventions in professional environments.

More recently, the EU Artificial Intelligence Act has expanded this framework. Of particular significance is Article 5, which prohibits the use of AI systems designed to infer emotional states in employment and education. This prohibition acknowledges the heightened risks of affective computing and emotion-recognition tools in contexts shaped by structural power imbalances and reduced individual autonomy. Collectively, these instruments provide a set of protections that, although uneven and still evolving, signal a growing regional commitment to regulating neurotechnologies in ways that align with fundamental rights and take account of workplace-specific vulnerabilities.

4.1. The role of Article 88 GDPR in workplace personal data processing

A number of GDPR provisions touch upon data processing in the workplace.[83] Central to this discussion, however, is Article 88(1), the so-called "opening clause", which empowers Member States to adopt specific rules for handling workers' data. This focus reflects the need to examine the scope of national discretion in an era where neurotechnologies allow unprecedented monitoring of emotions and

[83] See for instance, Article 24 GDPR on the duties of data controllers.

cognitive capacities, thereby raising concerns for dignity, autonomy, and mental privacy. The analysis proceeds from a human rights perspective, considering both the GDPR's text, the jurisprudence of the Court of Justice of the European Union (CJEU) and the European Court of Human Rights (ECtHR).

During the negotiations of the GDPR, efforts by the European Commission and the Directorate-General for Employment to establish a uniform regulatory framework for workplace data processing were ultimately resisted by Member States.[84] Article 88(1) was introduced as a compromise, granting flexibility to legislate in this sensitive field, often privileging consent as a legal ground for the processing of personal data in the employment context. Germany and Finland have enacted dedicated legislation under the clause, Spain and Greece have embedded workplace rules in broader data protection laws, while other jurisdictions, such as Croatia, France, and Luxembourg, have regulated specific issues, including biometrics and surveillance.[85]

Yet this discretion under this provision is not unlimited. Article 88(2) stipulates that national measures must include "suitable and specific safeguards" for dignity, legitimate interests, and fundamental rights. A 2022 preliminary ruling by the CJEU confirmed that national autonomy is conditional upon these safeguards.[86] The Court's reasoning in *Omega*, upholding a German prohibition on a laserdrome game on the basis of dignity, demonstrates that although conceptions of dignity may vary, they cannot undermine the core of EU law.[87] This logic mirrors Article 88 itself: diversity of national regulation is permissible only insofar as it grants full protection of dignity and fundamental rights.

The ECtHR's jurisprudence on Article 8 ECHR adds further guidance.[88] In *Barbulescu v. Romania*, the Court laid down criteria for evaluating workplace surveillance: prior notice, the scope and intrusiveness of monitoring, the existence of legitimate grounds, the possibility of less invasive alternatives, the impact on the employee, the presence of safeguards, and access to remedies.[89] These principles were extended to video surveillance in *Lopez Ribalda v. Spain.*[90] In *Surikov v. Ukraine*, a violation was found where an employer retained and shared mental health data beyond what was necessary for assessing promotion. The Court underlined that even where employers pursue legitimate interests, the collection of sensitive data must remain proportionate, lawful, and necessary in a democratic society.[91]

The question of consent has also been central in the Strasbourg Court's jurisprudence. In *Antović and Mirković v. Montenegro*, tacit consent to workplace video monitoring did not eliminate the intrusive character of surveillance,[92] while in *M.S. v. Sweden*, the Court stressed that prior consent is not determinative in assessing whether private life has been interfered with.[93] National practice is far from

[84] H. ABRAHA, *A pragmatic compromise? The role of Article 88 GDPR in upholding privacy in the workplace*, cit., 280.
[85] *Ivi*, 281.
[86] CJEU, Case C-34/21, *Bundesrepublik Deutschland v. JW*, Judgment, 30 June 2022, para 21.
[87] CJEU, Case C-36/02, *Omega Spielhallen – und Automatenaufstellungs – GmbH v. Oberbürgermeisterin der Bundesstadt Bonn*, Judgment, 14 October 2004, paras. 32-39.
[88] For a detailed analysis of the ECtHR's case-law on technological surveillance see A. SARTORI, *Il controllo tecnologico dei lavoratori,* cit., 14 ff.
[89] ECtHR, *Bărbulescu v. Romania* [GC], Application no. 61496/08, Judgment, 5 September 2017, para. 121
[90] CtHR, *López Ribalda and Others v. Spain* [GC], Applications nos. 1874/13 and 8567/13, Judgment, 17 October 2019, para. 116.
[91] ECtHR, *Surikov v. Ukraine*, Application no. 42788/06, Judgment, 26 January 2017, para. 94.
[92] ECtHR, *Antović and Mirković v. Montenegro*, Application no. 70838/13, Judgment, 28 November 2017, para. 44.
[93] ECtHR, *M. S. v. Sweden*, Application no. 20837/92, Judgment, 27 August 1997, paras 31, 35.

uniform: for instance, while in Portugal, consent cannot justify processing where employers derive a benefit[94] in Italy, reliance on consent for sensitive employee data is discouraged through soft law guidance.[95] Such divergences expose the lack of harmonisation across the EU and question the reliability of consent as a safeguard in relationships defined by power imbalance.

4.2. The EU Medical Device Regulation 2017/745

While a detailed examination of brain-altering neurostimulation devices lies beyond the scope of this paper, it has been highlighted that the use of commercially available neurotechnologies, marketed to enhance concentration, induce relaxation, or improve cognitive performance, raises serious concerns for the protection of mental integrity.[96] By directly modulating neural activity, such devices move beyond observation or inference, entering the domain of active alteration of mental states. Yet, existing regulatory regimes for neurostimulation devices, especially those available to consumers, remain largely focused on health and safety, paying limited attention to the broader implications of unconsented interference with neural functioning.[97]

The EU Medical Devices Regulation (MDR) 2017/745 already provides a legal framework to address some of these concerns. Binding across all Member States, the MDR includes provisions for so-called "non-medical purpose" devices, listed in Annex XVI, which are functionally analogous to medical devices but are marketed for general use, such as cognitive enhancement, improving stress levels or for wellness purposes. Neurostimulation devices are included in this category and are defined as equipment "intended for brain stimulation that apply electrical currents or magnetic or electromagnetic fields that penetrate the cranium to modify neuronal activity in the brain".[98]

The MDR requires manufacturers and other economic operators to comply with general safety and performance requirements set out in Annex I, including pre-market conformity assessments laid down in Annexes IX to XI. It is important to note that in December 2022, the European Commission issued *Common Specifications*[99] that further clarify obligations for manufacturers of non-medical, non-invasive neurostimulation devices, placed in class III, the highest risk category. Among these is the obligation to establish a comprehensive risk-management process to identify, minimise, or eliminate risks associated with neurostimulation technologies. These regulatory requirements, which have received critiques from

[94] Art 28(3) of Portuguese Law nº 58/2019 of August 8.

[95] Art 1.4.1.d of Italian Order 146/2019.

[96] For an in-depth consideration of the risks arising from brain-altering neurotechnologies see M. Sosa Navarro, S. Dura-Bernal, *Human Rights Systems of Protection From Neurotechnologies,* cit.

[97] A. Wexler, *A Pragmatic Analysis of the Regulation of Consumer Transcranial Direct Current Stimulation (TDCS) Devices in the United States*, cit., 672.

[98] Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC [2017], OJ L 117/1, para. 173.

[99] Commission Implementing Regulation (EU) 2022/2346 of 1 December 2022 laying down common specifications for the groups of products without an intended medical purpose listed in Annex XVI to Regulation (EU) 2017/745 of the European Parliament and of the Council [2022] OJ L 311/57; Commission Implementing Regulation (EU) 2022/2347 of 1 December 2022 laying down rules for the application of Regulation (EU) 2017/745 of the European Parliament and of the Council as regards reclassification of groups of certain active products without an intended medical purpose [2022] OJ L 311/72.

the neuroscience research community,[100] gain heightened importance in occupational settings, where workers may be subjected to subtle or explicit pressure to adopt neurostimulation tools aimed at enhancing productivity or attention. However, their effective implementation may be hindered by the conceptual ambiguity found in the wording of the risks that manufacturers are required to identify and mitigate, such as "atypical or other idiosyncratic effects" (art. 3.3.h, Annex VII) and "neural and neuro-toxicity risks" (art. 3.3. b, Annex VII), thereby limiting the overall effectiveness of the regulatory framework.

4.3. The AI Act and the regulation of emotion-inference technologies in the workplace

In May 2024, the European Union adopted the AI Act,[101] the first comprehensive and binding framework on artificial intelligence worldwide. Designed to act as a regulatory benchmark, the Act establishes a stratified risk-based system and entrusts oversight to the newly established AI Office and AI Board, which are tasked with issuing further interpretative guidance. As is often the case with ambitious legislation, several notions remain undefined, leaving space for legal uncertainty that will ultimately require judicial clarification.

For employment contexts, Article 5(1)(f) is of particular importance. It prohibits the marketing, deployment, or use of AI systems designed to infer human emotions in the workplace and in educational settings. Under Article 3(39), an emotion-recognition system is defined as "an AI system for the purpose of identifying or inferring emotions or intentions of natural persons on the basis of their biometric data". Recital 44 further elaborates that the term "emotion" covers a broad range of internal states, including happiness, sadness, anger, disgust, surprise, shame, contempt, satisfaction, and amusement. The prohibition is justified on the grounds of the structural imbalance of power in these environments and the resulting risks of discrimination and other harms. An exception exists for systems used strictly for medical or safety purposes, a carve-out that requires careful interpretation to avoid creating regulatory loopholes.

The Commission's interpretative guidelines confirm that the prohibition applies to a wide array of practices in employment, including the use of brain-reading neurotechnologies such as EEG or neuroimaging to "monitor emotions or boredom of employees" or to implement "well-being applications for making workers happier".[102] The guidelines further recall, in line with Recital 18, that physical states such as pain and fatigue do not fall within the notion of "emotion". Thus, while AI systems designed to detect driver fatigue and issue alerts are explicitly excluded, the broader exclusion of fatigue and pain is not confined to safety contexts. As a result, even in workplaces where fatigue is not a safety-critical factor, data of this kind would fall outside the scope of the prohibition set out in

[100] C. BUBLITZ, S. LIGTHART, *The new regulation of non-medical neurotechnologies in the European Union*, cit., 11-12.

[101] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 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) [2024] OJ L, 445/1.

[102] European Commission, *Guidelines on Prohibited AI Practices under the AI Act*, European Commission, 25 April 2025, p.71.

Article 5 and could therefore qualify as information that employers may collect and use in their decision-making processes, giving rise to discrimination risks based on neural-data inferences.[103]

The guidelines also clarify the breadth of the "workplace" concept, which should be understood as covering "any specific physical or virtual space where natural persons engage in tasks and responsibilities" assigned by an employer or organisation. This extends to employees, contractors, volunteers, and the self-employed. Significantly, Article 5(1)(f) also applies during recruitment processes, given both the imbalance of power and the intrusive nature of emotion recognition at the hiring stage.[104]

Moreover, Article 6(2) and Annex III classify several workplace-related AI systems (such as those used for recruitment, promotion, dismissal, or performance monitoring) as "high-risk", subject to stringent compliance requirements including risk management, human oversight, record-keeping, and public registration. Non-compliance can result in fines of up to €35 million or 7% of global annual turnover. Although the AI Act does not expressly regulate neurotechnologies, scholars such as Bublitz and Molnár-Gábor argue that its scope is wide enough to include technologies capable of inferring emotions or intentions from brain activity.[105] Recital 57 strengthens this reading by establishing a conceptual link to the GDPR (and therefore to the protection of personal data, including neural data) and underscoring the need for consistency in fundamental rights protections.

From the perspective of mental privacy and freedom of thought, two elements of the AI Act will be particularly decisive for workplace applications. First, the categorical exclusion of fatigue monitoring from the definition of emotion under Article 5(1)(f) risks leaving significant aspects of workers' neural data (which, as this paper has discussed, is critical for shielding against exploitative, productivity-driven neurosurveillance) outside the scope of protection. Second, the medical and safety exception to the prohibition on emotion recognition may invite expansive interpretations by employers, thereby weakening the ban's protective function. Together, these interpretive challenges will play a central role in determining whether the AI Act can provide meaningful safeguards for workers' mental privacy in practice.

5. Conclusive remarks

This paper has argued that the proliferation of neurotechnologies in the workplace raises a complex set of challenges for international law. A conceptual distinction between brain-reading devices, which collect and process neural data and engage primarily the rights to privacy and freedom of thought, and brain-altering devices, which directly modulate neural activity and thus implicate mental integrity, helps to illuminate distinct risk profiles. Yet the boundary is porous: intensive neuromonitoring can foster self-censorship and erode autonomy, while brain-altering techniques may suppress thought formation itself.

[103] *Ivi,* 79 and 84.
[104] *Ivi*, 81.
[105] C. BUBLITZ *et al.*, *Implications of the Novel EU AI Act for Neurotechnologies*, in *Neuron 112*, 18, 2024, 3014-3015.

In both cases, the foundational value of human dignity emerges as the decisive normative anchor, transcending debates over which single rights provides the most appropriate legal fit.[106]

Against this normative backdrop, the analysis has mapped an evolving but fragmented regulatory landscape. At the international level, Article 17 ICCPR, as elaborated by the Special Rapporteur on privacy, provides a principled framework for personal data processing that can be applied to neural data. In Europe, three instruments are of particular relevance: the GDPR, whose Article 88 delegates workplace-specific rules to Member States, producing uneven reliance on consent in an inherently imbalanced relationship; the MDR, which regulates non-medical neurostimulation devices but excludes brain-reading technologies, leaving them subject only to horizontal product-safety law; and the AI Act, which prohibits emotion-recognition systems in employment settings yet carves out medical and safety exceptions and excludes fatigue and pain from its scope. Considered together, these frameworks create overlapping protections but also inconsistencies and loopholes, exposing workers to protection gaps while at the same time generating regulatory uncertainty for innovation.

The precautionary principle, which has crystallised into customary international law, offers a valuable compass for navigating such uncertainty. Applied to neurotechnologies, it legitimises precautionary action in the face of scientific indeterminacy and underscores the need for a normative definition of 'thought'.

Soft law also plays a critical role in bridging regulatory gaps. The ILO's Code of Practice on Workers' Personal Data and the Council of Europe's Recommendation CM/Rec (2015)5, though formally non-binding, have shaped national laws, workplace policies and collective agreements with regard to this matter. Their practical influence illustrates why scholars regard soft law as an appropriate modality in contentious and fragmented fields such as workers' privacy, avoiding the paralysis of treaty negotiations.[107] Some even argue that sustained reliance on these instruments may contribute to the gradual crystallisation of customary norms.[108]

The European Union's approach further demonstrates the potential benefits of anticipatory regulation. By adopting the AI Act pre-emptively, the EU has sought to avoid the regulatory failures that allowed ungoverned digital platforms to entrench socially harmful practices. Although not free of interpretive ambiguities, this anticipatory model provides a useful template for addressing the disruptive potential of workplace neurotechnologies before harmful practices become entrenched.

In conclusion, the governance of workplace neurotechnologies requires an approach that is principled, precautionary, and anticipatory. International law must integrate the rights to privacy, freedom of thought, bodily and mental integrity, and non-discrimination into a coherent framework responsive to both neurosurveillance and neuromodulation. This entails interpretive development of existing instruments, reliance on soft-law guidance, and recourse to the precautionary principle to address uncertainty. Above all, regulation must ensure that technological innovation does not come at the

[106] F. D'AGOSTINO, *Bioetica e dignità dell'essere umano*, in C.M. MAZZONI (ed.), *Un quadro europeo per la bioetica*, Firenze, 1998, 153; G. PECES BARBA, *Dignidad humana*, in J.J. TAMAYO (ed.), *10 palabras clave sobre derechos humanos*, Navarra, 2005, 55.

[107] S. SIMITIS, *Reconsidering the Premises of Labour Law: Prolegomena to an EU Regulation on the Protection of Employees' Personal Data*, in *ELJ*, 1999, 50-51.

[108] A. SITZIA, *Lavoro e privacy: adempimenti obbligatori e procedure*, 2nd edition, Milan, 2012, 128-129.

expense of human dignity, and that the workplace remains a space for human flourishing rather than one of mental surveillance and cognitive exploitation.

Finally, legal frameworks must address not only the rights of individual workers but also the wider societal consequences of normalising continuous neuromonitoring, including its effects on deliberation, dissent, and creativity. Embedding multistakeholder and anticipatory governance upstream is essential to align regulatory standards with social values before technological dependence takes root.

Essays

Artificial Intelligence and Credit Scoring: The European Court of Justice Takes Action

Francesca Mattassoglio*

ABSTRACT: Recently, the European Court of Justice ruled on the subject of credit scoring calculations with techniques based upon artificial intelligence, which constitute cases of great interest, because the judge finally turned his attention to companies such as Schufa and B&D. This kind of dispute, between evaluated parties and scoring companies, will be probably destined to grow over the next few years and will almost certainly lead to questioning the rights of individuals with reference to the breadth of the right to information on the logic of the algorithm involved to achieve the final result.

KEYWORDS: artificial intelligence; credit scoring; proper to information; explainability; algorithm

SUMMARY: 1. Some introductory considerations – 2. The procedural events – 3. The general question relating to interpreting the concept of 'decision' contained in Article 22 GDPR – 4. Further specifications regarding the right of access recognised to the personal data owner, in the case of an automated creditworthiness assessment procedure – 5. The impact of decisions on companies acting as Schufa and B&D – 6. The provisions contained in the AI Act – 7. A nod to the new Directive 2023/2225/EU on consumer credit agreements – 8. Some brief conclusions.

1. Some introductory considerations

Recently, the European Court of Justice had the opportunity to rule on the subject of credit scoring calculations with techniques that resort to the aid of artificial intelligence,[1] with two

* *Francesca Mattassoglio: University of Milan-Bicocca. Mail: francesca.mattassoglio@unimib.it.* *This article was subject to a blind peer review process.*

[1] As regards the Italian doctrine on the subject of credit scoring, see, P. GAGGERO, C.A. VALENZA, *Le moderne tecniche di* credit scoring *tra GDPR, disciplina di settore e* AI Act, in *Rivista di diritto bancario*, 2024, 825 ss.; M. RABITTI, *Credit scoring* via *machine learning e prestito responsabile*, in *Rivista di diritto bancario*, 2023, 175 ss.; L. AMMANNATI, G.L. GRECO, *Piattaforme digitali, algoritmi e* big data*: il caso del* credit scoring, in *Rivista trimestrale di diritto dell'economia*, 2021, 305; F. MATTASSOGLIO, *La valutazione 'innovativa' del merito creditizio del consumatore e le sfide per il regolatore***,** in *Diritto della banca e del mercato finanziario,* 2, 2020, 187-220; ID., *Innovazione tecnologica e valutazione del merito creditizio del consumatore,* Milano, 2018, 9 ff.

different rulings, which constitute cases of great interest, because the judge has finally turned his attention to companies such as Schufa and B&D.[2]

These entities base their business model on the exploitation of data, selling, to third parties, assessments about the creditworthiness of any potential user of services, on German and Austrian territory respectively, without, however, ever having been forced – at least so far – to comply with the provisions contained in Article 22 of the Regulation on the Protection of Personal Data (GDPR), i.e., the only ones that exist to date regarding these automated procedures. The assumption, behind which said companies have always taken refuge, was based on the fact that they were mere service providers, in a broader supply chain, which then did not see them in any way involved in the final decision, capable of affecting the legal situation of the user, that is, for example, in the provision of credit.

As is well known, the use of automated data processing systems has been the subject of specific attention by the regulator, only recently, with the now famous regulation on artificial intelligence (known as the AI Act) [3] – which brought credit scoring within the scope of those deemed high-risk activities, which require compliance with strengthened prescriptions.

We will return to this aspect in more detail later. Still, here it is appropriate to underline how, from now on, credit scoring companies, which use AI for their evaluation activity, will be subjected to a more rigorous regulatory framework, given that they will have to comply with both the provisions contained in the GDPR and those of the AI Act, given that the regulation on artificial intelligence itself is without prejudice to data protection legislation (recital 10 of the AI act).

Precisely as confirmation of this growing attention, reference must be made to the conclusions of the Advocate General, in the second decision, which, starting from the Schufa case, addressed the issue relating to the interpretation of the notion of "significant information on the logic used" in the context of an automated decision-making process, pursuant to art. 15, par. 1, letter (h) GDPR, with particular attention to the balance between the right of access to information, on the one hand, and the protection of rights, trade secrets included therein, on the other.

Both sentences confirm that, at the European level, a new line of jurisprudence is taking hold, destined, at least we hope, to have a profound impact on the sector.

In particular, from a legal point of view and according to a value we could define as general, the Court's rulings are progressively broadening and better delineating the scope of application of the art. 22 GDPR, far beyond the traditional notion of 'decision', and regardless of the scope in which that automated process is used.

---

[2] European Court of Justice, judgment of 7 December 2023, Case C-634/21 and 27 February 2025, Case C-203/22. For comments, see F. CIRAOLO, *Le valutazioni automatizzate del merito creditizio nel quadro regolatorio europeo. Quale futuro per il credit scoring algoritmico?,* in *Rivista di Diritto bancario*, 2025, 105 ss.; A.G. GRASSO, *Decisioni automatizzate e merito creditizio: la Corte di giustizia sul credit-scoring*, in *Banca borsa e titoli di credito*, 77, 2024, 730 ss.; CIPL, *Decoding Responsibility in the Era of Automated Decisions: Understanding the Implications of the CJEU's SCHUFA Judgment*, October 2024, in https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_decoding_responsibility_automated_decision_making_oct24.pdf; A. AZA, *Scores as Decisions? Article 22 GDPR and the Judgment of the CJEU in SCHUFA Holding (Scoring) in the Labour Context*, in *Industrial Law Journal,* 53(4), 2024.

[3] Reg. UE/1689/2024 which establishes harmonized rules on artificial intelligence and modifies the regulations (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 and (UE) 2019/2144 and directives 2014/90/UE, (UE) 2016/797 and (UE) 2020/1828 (AI act).

Secondly, and from a more sectoral point of view, precisely thanks to a broader interpretation of the concept of decision, the European court appears to be oriented towards extending to companies such as Schufa and B&D, not only the provisions contained in Article 22, but also those relating to the disclosure requirements of Article. 15.
A significant evolution, given the weight that the decisions of these subjects have assumed in our societies.

2. The procedural events

Before moving on to consider the more strictly legal aspects, it is necessary to recall, albeit briefly, the events from which the appeals began: the first presented by a German citizen who had been denied credit by a financial intermediary, in the face of a negative credit scoring provided by Schufa; the second, however, initiated by an Austrian user who, due again to a negative evaluation conducted this time by B&D, had not been able to obtain the extension of a telephone contract, for a truly negligible value such as 10 euros per month.
In Germany, Schufa Holding AG (acronym for Schutzgemeinschaft für allgemeine Kreditsicherung) is a private company which basically has a monopoly concerning the assessment of the creditworthiness of persons entering German territory, thanks to over 9,000 contractual partners, which include telecommunications and electricity service providers, banks, savings banks, cooperative credit banks, and credit card companies.[4]
Based on the collected data, Schufa generates the so-called scoring of the subject considered, which constitutes an estimate of the subject's creditworthiness, expressed as a percentage, with 50% indicating a significant risk. In comparison, above 97.5% means a minimum risk.
That kind of score, among other things, not only interferes with access to credit, but also with the search for an apartment, with access to fundamental services such as electricity or telephone services, given that the conclusion of each of these contracts cannot ignore a verification of the position of the individual by Schufa.
In the case brought to the attention of the European Court, a German woman was thus denied a loan grant, by an intermediary, based on a negative opinion prepared by Schufa who, subsequently, refused not only to communicate to the interested party the precise data that had been placed as the basis of such an assessment, but also to correct any errors contained therein.
In fact, the company limited itself to providing a summary answer, invoking commercial secrecy to withhold more precise information about the calculation method. In addition to this statement, Schufa's defence was based mainly on its contention that it was not responsible for the infringement of the applicant's legal rights, given that it simply provided a report to the intermediary, who then took the final decision to deny credit.
Given such a position, the applicant requested the intervention of the German data protection authority, which quickly endorsed the agency's position, stating that the entity had not infringed the applicable rules. And this is why the matter then came before the European judge.

[4] https://www.schufa.de/en/ (last visited 29/09/2025).

In the second case, on the other hand, a citizen who was Austrian this time had been denied the extension of a contract by a mobile operator, which would have resulted in a monthly payment of only EUR 10, because of an assessment of financial unreliability, conducted in an automated manner by the company B&D, which was again a company specialising in the provision of such evaluations, like its German counterpart.

In this case, however, the national data protection authority accepted the petition submitted. It obliged the company to provide relevant information on the logic used in automated decision-making. However, it was then that B&D challenged the decision and, subsequently, having its appeal rejected, only partially enforced it.

The matter thus finally reached the European Court since, according to the appellant, B&D did not adequately fulfil its information obligations, producing untruthful information to such an extent that, based on the explanations offered, the applicant's creditworthiness would even have to be particularly high. Therefore, in apparent contradiction with the final evaluation judgment rendered.

Having said this, from a factual point of view, it is now possible to move on to considering the more strictly legal profiles of the events in question.

3. The general question relating to interpreting the concept of 'decision' contained in Article 22 GDPR

As can be seen from the facts just reported, the main question, which the European court had to address first of all, concerns Article 22 (1) of the GPR and, in particular, whether that rule must be interpreted as constituting an "automated decision-making process relating to natural person" within the meaning of that provision, automated calculation, by a company providing commercial information, of a probability rate based on personal data relating to a person and concerning the latter's ability to honor payment commitments in the future, if the stipulation, execution or termination of a contractual relationship with that person by a third party, to whom such probability rate is communicated, decisively depends on that probability rate.

In fact, only in the case of an automated decision-making process does the rule allow a whole series of requirements and protections to be triggered for the benefit of the interested party, first of all, his right not to be subjected to a decision based solely on automated processing, including so-called profiling, if it produces legal effects or similarly significantly affects his person.

As is known, there are three conditions that, cumulatively, must exist for the rule to be applicable: first, there must be a 'decision'; secondly, this decision must be based "solely on automated processing, including profiling"; finally, that decision must be able to produce "legal effects" or in any case intervene "similarly significantly" on the person concerned.

In light of these prescriptions, the question concerns the perimeter of the notion of 'decision' and the consequent need to adapt to a process that increasingly uses advanced technologies. In particular, from this point of view, it must be taken into account that the advent of AI systems has not only allowed the processing of enormous quantities of data at speeds unimaginable for human beings but also, and at the same time, increased the number of subjects potentially involved in the decision-making process, making it more complex.

It is no coincidence that the AI act itself recalls the existence of a real "value chain" [5] linked to AI, which often involves the co-presence of a plurality of subjects who may, from time to time, be engaged in it.
From this, one of the undoubtedly most challenging elements that this type of technology poses to the regulator is identifying the responsible agent for specific conduct.
In this new context, the final 'decision' can often intervene only upstream of a long and varied chain of events, each of which can constitute its fundamental prerequisite. Thinking then that only the last step, the one sometimes now devoid of any actual decision-making value – understood as a possible choice between a plurality of alternatives – is the one detrimental to the position of the interested parties, would be inappropriate.
Indeed, the matter under comment demonstrates that the actual infringement of the loan applicant's right/interest had already been realised well before the inevitable and subsequent 'formal' denial of the loan by the financial intermediary, namely, at the moment when the rating company rendered its judgment.
In light of this awareness, according to the court, it would henceforth be necessary to recognise a broad scope of the concept of 'decision' in Article 22 in question, also by virtue of the wording of recital 71, such that it could also include a mere "measure" capable of producing "legal effects concerning him" or of affecting "in a similar way significantly his person", which can also be, precisely, an automatic refusal of an online credit request or electronic hiring practices, if there has been no human intervention.
Consequently, the decision should also include the 'mere' result of calculating a person's solvency in the form of a probability rate relating to that person's ability to honour payment commitments in the future.
By virtue of this assumption, the judge's argument becomes much simpler. Clarified that even a judgment such as that expressed by Schufa can be understood as an automated decision, pursuant to Art. 22, there are no longer any obstacles to identifying the presence of the other two conditions required by the rule.
As regards the second condition, in fact, the Advocate General, in paragraph 33 of his Opinion, had also considered it common ground that an activity such as that of Schufa met the definition of "profiling" in Article 4 (4) of the GDPR.
Finally, with respect to the effects, according to the judge, the probability rate, given its weight in the context of the credit-granting decision, can significantly impact the interested party.
Consequently, given that the probability rate, established by a company providing commercial information and communicated to a bank, plays a decisive role in granting credit, the calculation of this rate must be classified in itself as a decision it produces in relation to a data subject "legal effects

[5] This concept is referred to Cons. 9 of the AI act, while Art. 16 provides that "any distributor, importer, deployer or other third party is considered a supplier of a high-risk AI system for the purposes of this Regulation and is subject to the obligations, according to the principle of liability along the AI value chain". For the first comments regarding the AI Act, see, C. CASONATO, B. MARCHETTI, *Prime osservazioni sulla proposta di regolamento dell'Unione Europea in materia di intelligenza artificiale*, in *BioLaw Journal*, 2021, 415 ff.; G. FINOCCHIARO, *La proposta di regolamento sull'intelligenza artificiale: il modello europeo basato sulla gestione del rischio*, in *Diritto dell'informatica*, 2022, 303 ff.; V. LEMMA, *Intelligenza artificiale e sistemi di controllo: quali prospettive regolamentari?*, in *Rivista trimestrale di diritto dell'economia*, 3, 2021, 319 ff.

concerning him or her or similarly significantly affecting him or her", within the meaning of Article 22 (1) of the GDPR.
This statement undoubtedly subjects companies that conduct scoring activities to the regulations on the protection of personal data, specifically those relating to automated processing.

4. Further specifications regarding the right of access recognised to the personal data owner, in the case of an automated creditworthiness assessment procedure

The clear statement of the position of the European court, regarding the obligation on scoring companies to comply with the principles set out in Article 22, could only open Pandora's box of appeals against similar parties and require further details, especially about the extent of the rights of individuals, subjected to such procedures around Europe.
In the case raised by the Austrian consumer, for example, the question concerned the interpretation of Article 15 and, more specifically, the extent of the right of access enshrined therein, especially in the case of balancing with Article 4 (6), which protects the commercial or business secrecy of the controller or third parties.
As the Advocate General himself specifies, starting precisely from the Schufa decision – which constitutes the logical prerequisite for the dispute – this case requests to integrate its content to the point of identifying how to resolve any conflict between interests involved and, again, establish what breadth those "significant information on the logic used" must take in the context of an automated decision-making process, pursuant to art. 15, par. 1, letter. (h), GDPR.
And it is in this regard, starting from point 40 of the decision, that the Court undertakes a complex reasoning aimed at reconstructing the meaning of the phrase "significant information on the logic involved", resorting – even – to the semantic linguistic apparatus of the different languages of the EU countries, in search of sufficiently precise concepts to adapt to the multifaceted reality of automated decision-making processes.
Here, it is useful, in particular, to recall the passage where the judge specifies that in the French version 'informations utilis' it would take on a connotation relating to 'functionality' which it would also retain in the Dutch 'nuttige', or in the Portuguese 'úteis'.
Romanian, however, would be more expressed in the sense of the relevance ('relevant') of the information to be provided. At the same time, in other translations their importance ('significant' in Spanish and 'istotne' in Polish) would be more accentuated. Finally, either the term 'aussagekräftig' used in German or the 'meaningful' in English would seem to refer more to the 'good intelligibility' and 'quality' of the information used for the explanation.
In point 42 of the decision, the "logic involved", again with reference to the automated decision-making process, according to the judge, should cover "a wide range of 'logics' of use of personal data and other data to obtain, with automated means, a certain result". In support of such a conclusion, some examples of translation into Czech and Polish are thus recalled, where the terms 'postupu' and 'zasady' could be translated with a meaning equivalent to 'procedure' and 'principles'.
These considerations enable the Court to state that, under the legislation thus interpreted, it would be possible to grant the data subject the right to "verify that personal data concerning him or her are

correct and processed lawfully" – that is, it would be, first of all, the data controller must provide the applicant with a "full" and "faithful" copy of the personal data that have been used as part of the automated procedure, to allow him to exercise his rights.

In other words, this would be the right to rectification (art. 16), to erasure of data ('right to be forgotten') (art. 17), to restrict processing (art. 18), as well as to object to processing (art. 21) and to take legal action (art. 79 and 82) which could never find any application in the absence of this initial passage of information.

Greater attention should be paid if the subject has been profiled.

But what should this information be like to be considered adequate in the case of an automated process?

To resolve the question, the court refers to the provisions of Article 12 (1) that "information intended for the person concerned must be concise, easily accessible and easy to understand, and formulated in simple and clear language", in other words, 'intelligible' for the addressee, unless it loses any usefulness. And this is where the issue becomes crucial.

In fact, in automated procedures, the data underlying the evaluation process can be pervasive and derived from other data (i.e., inferred or induced). Hence, the data subject must also be made aware 'of the context' in which this information was used.

This would, therefore, result in the recognition of a far broader right of access than the traditional application of Article 15 in conjunction with the previous Article 12, by virtue of the special features of the automated process.

Therefore, the "significant information on the logic involved" should relate to "any relevant information relating to the procedure and principles of use of personal data to obtain, by automated means, a certain result". In contrast, the obligation of transparency would require "that such information be provided in a concise, transparent, understandable and easily accessible form" (par. 50).

Consequently, the data subject should have "a genuine right to obtain explanations regarding the functioning of the mechanism underlying an automated decision-making process of which that data subject has been the subject and the result to which that decision has led".

The conclusion of this reasoning – or rather, the very emphasis given to the element of comprehensibility –, however, prompts the judge to exclude the algorithm from the list of information that must be disclosed by the data controller, by reason of its complexity and the consequent presumed incomprehensibility for the individual.

By way of reference, the data controller would therefore not be under an obligation to disclose information that presents a high level of complexity, such as to render it incomprehensible, without adequate technical expertise, thereby making it possible to exclude from communication the very algorithms used in the context of automated decision-making. In fact, according to the Court, "the simple communication of a complex mathematical formula, such as an algorithm, or the detailed description of all the stages of an automated decision-making process" can be considered to be in line with the fulfilment of said obligation. Otherwise, it would be necessary for the controller to describe "the procedure and principles actually applied", to allow the data subject to "understand which of his or her personal data have been used in what way in the automated decision-making process in question,

without the complexity of the operations to be carried out in the context of the automated decision-making process exempting the data controller from his duty of explanation" (par. 61).

It follows from this that, in the case of a credit scoring assessment, it could, for example, be deemed sufficient "to inform the data subject how a change at the level of the personal data taken into account would have led to a different result" (par. 62).

As regards the profile relating to the comparison of interests (both of third parties and relating to industrial secrecy), the Court refers to the supervisory authority or the competent court the task of concretely weighing the interests at stake and establishing the scope of the right of access, which could be recognised to any applicant.

In light of what has been stated so far, it is clear that these conclusions only partially address the appellant's interests, leaving the most delicate issue unresolved: the algorithm used to conduct the evaluation.

The conclusion reached by the European judge, however, should not be surprising, especially given the complexity now recognised in decision-making processes that use artificial intelligence.[6]

In particular, and without naturally being able to go into the details of the issue, it has now been highlighted how three different causes of opacity often characterise AI:

(a) The first of an institutional nature, namely, linked to the presence of know-how and secrets maintained by the software developers. A type of darkness that can certainly be addressed and resolved, even by a solution such as that put forward by the Court of Justice, i.e. by leaving the task of balancing interests to a judge or a third-party authority;

(b) The second is of a technical nature, linked to the complexity and difficulty for those who do not possess specific IT/mathematical skills. A type of opacity that determines, in fact, the potential distinction of the population between ignorant masses and an elite of technicians capable of understanding its functioning;[7]

(c) and, finally, the most dangerous one by far: an epistemic obscurity that becomes the real problem. In this case, reference is made to the complex operation of automated decision-making systems, which are increasingly characterised by the absence of fundamental epistemically relevant elements that would provide an understandable explanation in human terms.

According to this setting, consequently, the behavior of neural networks although it can be mathematically precise – thanks to the presence of a possible computational explanation – it could, at

[6] With regard this concept, see B.D. MITTELSTADT, P. ALLO, M. TADDEO, S. WATCHTER, L. FLORIDI, *The Ethics of Algorithms: Mapping the Debate*, in *Big Data & Society*, 3, 2016, 2; S. WACHTER, B.D. MITTELSTADT, L. FLORIDI, *Why a Right to Explanation of Automated Decision-making Does Not Exist in the General Data Protection Regulation*, in *International Data Privacy Law,* 2, 2017, 76-99, which focuses on the problem relating to the explicability of AI and, on the other hand, positions such as that of G. WHEELER, *Machine Epistemology and Big Data*, in L. MCINTYRE, A. ROSENBURG (edited by), *Routledge Companion to Philosophy of Social Science*, New York, 2016, and that of C. GLYMOUR, *The Automation of Discovery*, in *Daedalus*, 2004, who are instead enthusiastic about the potential of machine learning techniques, to the point of considering the need to reach an explanation now completely obsolete.

[7] In this regard, see Y.N. HARARI, *Nexus: A Brief History of Information Networks from the Stone Age to AI*, London, 2024.

the same time, continue to be "epistemically opaque" from the point of view of its ability to make the decision "intelligible" and therefore then "shareable" by the human being.[8]

For these reasons, it is clear how the possibility of reasonably interpreting the content of Article 15 is most uncertain, and more specifically, what meaning is to be attributed to that right to obtain "significant information on the logic involved" when artificial intelligence is used. This is why some authors have suggested – even before the GDPR came into force – that the provision should be interpreted as giving the individual a mere right to 'be informed', ex ante, of the existence of an automated procedure and the logic underlying the final decision.[9] However, it would have been too complex to demand an ex post explanation of how that specific and single final decision was reached.

An extensive problem, which is becoming increasingly evident and current.

5. The impact of decisions on companies acting as Schufa and B&D

One of the most interesting aspects of the acts under comment is undoubtedly that of referring to the use of artificial intelligence[10] by entities such as Schufa and B&D, which, although performing a fundamental function in the lives of citizens, have so far been deemed outside the scope of Articles 22, 15 and 12 of the GDPR.

Given the role these entities play, many are even convinced that they are public, when in reality, as anticipated, they are private.

As noted above, this type of company falls within the scope of the so-called credit rating agencies, which can be public, if managed by the central bank, or private.[11]

In Italy, for example, both categories are present. While public risk centres were initially established for macroeconomic reasons, private ones were created to acquire greater knowledge of their potential counterparties. In this second case, the data is exclusively collected by entities that sign specific contractual agreements with one another, which, as we have seen, now range across the most diverse sectors. In this regard, it is worth reiterating that the scores these companies assign affect access to credit and can influence the possibility of renting a house, obtaining a connection to domestic utilities, etc.

[8] S. Zipoli Caiani, *A cosa pensano le macchine? Efficienza e opacità nelle reti neurali artificiali*, in AA.VV., *Filosofia dell'intelligenza artificiale*, Bologna, 2024, 21 ff.

[9] Moreover, significant interpretative doubts had already concerned the provision contained in the 1995 directive. On this topic, see D. Korff, *New Challenges to Data Protection Study – Working Paper No. 2: Data Protection Laws in the EU: The Difficulties in Meeting the Challenges Posed by Global Social and Technical Developments*, January 15, 2010, European Commission DG Justice, Freedom and Security Report, *online*.

[10] K. Langenbucher, *Consumer Credit in The Age of AI – Beyond Anti-Discrimination Law*, *SAFE Working Paper*, 2022, 369; C. O'neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, New York, 2016.

[11] Regarding this point, see L. Buonanno, *Un modello giuridico europeo di 'credit reporting industry'*, in *Banca borsa e titoli di credito*, 2022, 582 ff.; C. Venditti, *Il sistema binario italiano per la centralizzazione delle informazioni sui rischi creditizi*, in *Diritto del mercato assicurativo e finanziario*, 2021, 156 ff.; A. Sciarrone Alibrandi, F. Mattassoglio, *Le centrali dei rischi: problemi e prospettive*, in *Diritto della banca e del mercato finanziario*, 2017, 764-785.

From this point of view, having established the obligation to apply the relevant articles of the GDPR constitutes an essential step towards companies destined to have an increasingly relevant role in consumers' lives.

In Italy, these private credit rating agencies (so-called Società di Informazione Creditizia o SIC) are subject to a code of conduct on the subject ('Code of Conduct') (adopted by the Resolution of the Privacy Guarantor of 6 October 2022), whose Art. 10, which precisely with reference to automated procedures, states that "In cases where personal data contained in a SIC are also processed by automated scoring treatments or decision-making processes, the manager and the participants, it being understood that the managers do not take under the terms of the Regulation any decision that may affect the rights and freedoms of data subjects".

In this case, therefore, it is the same code that has consistently ruled out that the activity of managers could be brought within the scope of the notion of "decision", referred to in Article 22 of the GDPR.

Consequently, the ruling of the Court of Justice will certainly produce significant effects on the sector, especially since, in the meantime, the AI Act has also come into force, which is destined to impose much more penetrating obligations on information system managers than what has been foreseen so far.

6. The provisions contained in the AI Act

As already anticipated, the AI Act devotes particular attention to the assessment of the creditworthiness of consumers (natural persons), given that said judgments "determine the access of such persons to financial resources or to essential services such as housing, electricity and telecommunications services" (recital 58).[12] Due to this role, these automated systems could therefore lead "to discrimination between people or groups and can perpetuate historical models of discrimination, such as that based on racial or ethnic origin, gender, disabilities, age or sexual orientation, or can give rise to new forms of discriminatory impacts".

For this reason, the regulation brings this type of assessment back into the context of AI use cases considered high risk.[13]

Note that, in the opinion of the writer, it would be more absurd than ever to believe – as someone has done[14] – that the regulation should be interpreted restrictively, excluding from its scope precisely cases of credit scoring that are purely aimed at obtaining financial resources – think of the hypothesis of a mortgage – to refer instead exclusively to assessments connected with obtaining good and services deemed essential, emphasising the differences in terminology between recital 58 and the Annex below.

---

[12] L. Rocco Di Torrepadula, *L'uso del credit scoring algoritmico nella valutazione del merito creditizio*, in *Banca borsa e titoli di credito*, 78, 2025, 213; M. Rabitti, *Credit scoring* via *machine learning*, cit., 175 ff.

[13] In this respect, see the list in point 5 of Annex III dedicated to access to essential private services and essential public services and services and their use, where it provides that "IA systems intended to be used to assess the creditworthiness of natural persons or to establish their creditworthiness, with the exception of IA systems used for the purpose of detecting financial fraud". For a deeper analysis, see D. Di Sabato, G. Alfano, *L'impiego dell'IA per condizionare e valutare le persone tra limitazioni e divieti: qualche considerazione critica sulla proposta di Regolamento sull'IA elaborata dalla Commissione europea*, in *Rivista di diritto dell'impresa,* 2022, 281 ff.

[14] G. Spindler, *Algorithms, credit scoring, and the new proposals of the EU for an AI Act and on a Consumer Credit Directive*, cit., 243, regarding the interpretative doubts of credit scoring systems falling within the scope of high-risk systems, due to a mismatch between the provisions contained in recital 58 and Annex III.

Not only is it clear that financial resources are a fundamental requirement for an individual to access essential goods and services, such as housing, but credit scoring linked to financing requests is one of the thorniest sectors, in which individuals certainly deserve protection. It can therefore be defined as delicate and high-risk ex se.

A similar interpretation is also confirmed in the new directive on consumer credit 2023/2225, shown below.

Here, it is certainly not possible to analyse all the provisions in the AI Act in detail. Still, it is considered appropriate to focus on two aspects that highlight deep connections with the principles affirmed by the European Court of Justice.

First, European regulation pays much attention to considering the multiplicity of procedural phases that tools such as artificial intelligence may involve. In recital 53, for example, it introduces the distinction between cases where the use of AI is capable of materially influencing the decision-making process – that is, where the algorithmic component takes on a decisive role for the decision, consequently influencing interests substantially –, and hypotheses where no such impact can be recognised, and the final decision can be considered separately.

With reference to this second category, the first hypothesis mentioned is that in which AI is used to carry out a task defined as "restricted procedural", i.e., a mere transformation of unstructured data into structured data, as in the case in which it is used exclusively to catalogue a series of documents. In this case, the activity would not involve significant risks, given its limited nature.

The second hypothesis concerns, however, those tasks aimed at improving the results of "a previously completed human activity", in which AI would therefore only allow the addition of a further level, as in applications that enable improving the language used in writing.

Thirdly, the hypothesis of decision models or deviations from previous decision models is recalled: when AI is used after a human decision phase, to verify its validity ex post. Finally, reference is made to the use of AI in the context of a mere preparatory phase of a human decision, as in file management, indexing, research, textual and voice processing, linking data to other data sources, or systems used for translation.

Therefore, the new regulation seeks to provide greater clarity on complex decision-making processes and to distinguish the phases in which automation can intervene, thereby clarifying when the real decision-making moment should be placed, rather than in merely preparatory or subsequent phases.

Secondly, the AI act confirms the tendency to broaden the scope of potential recipients of obligations, taking into consideration a very varied audience of subjects involved in the process, which can lead to the adoption of the final decision, according to an approach perfectly in line with the evolution of the Court of Justice, which has just been taken into account.

Precisely for this reason, for example, Article 3 of the regulation within the scope of its definitions recalls various figures who must be considered involved in the process and therefore potentially subject to the obligations imposed, such as: suppliers, deployers, authorized representatives, importers, distributors, operators, each of whom could be called into question if he intervenes, in some way, on the artificial intelligence system.

These forecasts demonstrate how the regulation intends to consider all entities involved in using AI, so that complexity does not translate into irresponsibility.

From this perspective, apparent similarities can be seen between the jurisprudential and regulatory paths.[15]

7. A nod to the new Directive 2023/2225/EU on consumer credit agreements

Finally, the Directive 2023/2225/EU on consumer credit agreements (Consumer Credit Directive, CCD II), which repeals the previous Directive 2008/48/EC, and must be implemented by November 2025, introduces other essential provisions.[16]

In particular, for our purposes, the wording of Article 18 (3), which requires not only that "The assessment of creditworthiness be carried out based on relevant and accurate information on the income and expenditure of the consumer and on other information on the economic and financial situation, which is necessary and proportionate in relation to its nature, duration, is especially relevant to the value and risks of credit for the consumer", but goes so far as to exemplify some types of data that can be used such as "evidence of income or other sources of reimbursement, information on financial assets and liabilities or information on other commitments financial".[17]

Whereas the provision introduces an express prohibition on the use of sensitive data as referred to in Article 9 (1) of GDPR 2016/679 and requires the "relevance" of internal or external sources from which information may be collected by specifying that "Social networks are not considered to be an external source for this Directive".[18]

With specific reference to an assessment carried out by automated data processing, paragraph 8, of Article 18 then assures, to the consumer applying for the loan, the right to be able to demand and obtain human intervention, which takes the form of the possibility of arriving at a clear and comprehensible explanation of the procedure used for the assessment, including its logic and any risks; the right to express one's opinion and, finally, to request a review of the evaluation and decision relating to the credit application. In the event of a negative outcome, the right to a human review and to contest the decision is reiterated.

Therefore, once again, a series of guarantees and rights for subjects subjected to an automated evaluation procedure which, if on the one hand confirm the rulings of the judge and the advocate general in question, on the other end up returning once again to the problem relating to the interpretation of that right to obtain a clear and understandable explanation already mentioned.

---

[15] Regarding the relationship between GDPR and AI Act, see P. FALLETTA, A. MARSANO, *Intelligenza artificiale e protezione dei dati personali: il rapporto tra Regolamento europeo sull'intelligenza artificiale e GDPR*, in *Rivista italiana di informatica e diritto*, 2024, 119 ff.

[16] M. ORTINO, *La terza direttiva sul credito ai consumatori: distinzioni e complementarità nella tutela di interessi pubblici e privati*, in *Rivista della regolazione del mercato*, 2, 2024, 569.

[17] G. FALCONE, *Prime riflessioni sulla Direttiva CCD II: le informazioni e la valutazione del merito creditizio*, in *Rivista della banca e del mercato finanziario,* 4, 2024, 613; F. TRAPANI, *La nuova direttiva 2023/2225/UE sul credito al consumo: note in tema di educazione finanziaria, merito di credito e servizi di consulenza sul debito*, in *Le Nuove Leggi Civili Commentate,* 3, 2024, 754; P. GAGGERO (a cura di), *Primo commento sui criteri ordinatori della Direttiva UE n. 2023/2225 relativa ai contratti*, in *Rivista Trimestrale di Diritto dell'Economia*, suppl. al n. 4, 2024.

[18] N.M.F. FARAONE, *Spunti ricostruttivi in materia di profilazione e valutazione del merito creditizio nella nuova strategia europea dei dati*, in *Analisi giuridica dell'economia*, 44, 2025, 267 ff.

8. Some brief conclusions

Trying at this point to pull the strings of our reflections, we can note how the European legal system, thanks to the interpretative activity of the Court of Justice and the intervention of the regulator, has now decided to focus its attention on the activity conducted through the use of artificial intelligence, including the case in which companies operating in the field of credit scoring are involved, putting an end to a season of substantial regulatory uncertainty.

Indeed, probably, the events referred to here will constitute only the first signs of a dispute, between evaluated subjects and scoring companies, which will be destined to grow over the next few months and years and which will almost certainly lead to questioning and better examining the question of the rights of individuals, especially with reference to the breadth of the right to information on the logic of the algorithm that is used to reach the final judgement.

In the opinion of the writer, this will, in fact, be the front on which the most challenging game will be fought, given that the algorithm is the heart, the very essence of an automated procedure. The only one that justifies and determines the final result based on the specific set of data it was fed.

It is, therefore, by its nature complex and destined to become increasingly obscure as automated processes evolve and gain influence.

For this, it will probably be necessary to be aware of what has already been labelled as the conjecture on Floridi's "certainty-scope",[19] in other words, the attempt to mathematically prove the existence of a fundamental trade-off between the epistemic certainty relating to the result of a process that uses an AI system and the breadth of its operation and purpose. According to this hypothesis, only in systems that operate in restricted contexts is it possible to hypothesise complete epistemological certainty as their breadth increases; however, their inextricable unknowability becomes increasingly pronounced.

Conjecture on which, certainly in the coming years, the judge and jurists more generally will also have to begin to reflect.

[19] L. FLORIDI, *A Conjecture on a Fundamental Trade-Off between Certainty and Scope in Symbolic and Generative AI, in Philosophy & Technology,* 38, 2025, 93.

Essays

Digital Technology and the Responsibility of French Legal Professionals

Philippe Pierre*

ABSTRACT: The individual responsibility of legal professionals plays a key role in building trust with their clients. The increasing number of digital tools available to them is changing their daily practices, making many tasks simpler and quicker. However, this also raises questions about the legitimacy of their involvement, given that complementarity can also become competition. Applying the prism of civil liability to the duty to advise and the technical diligence expected of legal professionals provides a way of addressing this issue as it evolves.

KEYWORDS: responsibility; legal professionals; digital tools; duty to advice; technical diligences

SUMMARY: 1. Introduction – 2. The duty of legal professionals to provide advice within their digital environment – 2.1. Do digital tools change the standard of a diligent and competent practitioner? – 2.2. Are practitioners who use digital tools definitively liable? – 3. Technical diligence of legal professionals in their digital environment – 3.1. Does distancing documents alter liability terms? – 3.2. Could the legal professionals' liability protect them against the unbridled digitisation of their work? – 4. Conclusion.

1. Introduction

The digital environment of legal professions is multifaceted, incorporating document databases for AI use, jurimetrics, smart contracts, various blockchains and remote working for drafting and storing documents among other things. The advantages are well established: automation of repetitive and time-consuming tasks allows professionals to focus on more complex and strategic work, and improves efficiency in legal research, and even addiction to prediction. However, faced with this enthusiasm, which will undoubtedly be heightened by generational renewal, the commitment to responsibility may seem like a hindrance at first glance. Moreover, a LexisNexis survey found that 85% of French legal professionals already have ethical concerns, which are not far removed from Hans Jonas's 'principle of responsibility' developed in 1979. Jonas emphasises how "the promise of modern technology has been turned into a threat".

However, the issue of responsibility in relation to digital tools, which we will mainly study based on the French legal model, has yet to be resolved. In most cases, the use of digital tools will have no impact on the legal professional's commitment to responsibility, preserving both the paradigm of good professional practice and the protection of their clients. Nevertheless, in certain circumstances, the

* *Professor at the University of Rennes; Holder of the International Chair in Notarial Law. Mail: philippe.pierre@univ-rennes.fr.* *This article was subject to a blind peer review process.*

digital environment may extend this responsibility, although this extension is not necessarily detrimental to the practitioner. We will see this in the classic analysis of the two sources of liability for legal professionals: failure to provide advice (I) and legal errors (II).

Before addressing these issues through the prism of liability, we will briefly discuss insurance, without which our comments would risk becoming a circular argument. Currently, the collective insurance taken out by legal professionals – lawyers and notaries – appears flexible enough to 'absorb' the impact of digital technology. The model is effectively that of '*all risks except*', based not on a list of covered risks, which should specifically target digital risks, but on blanket cover provided that the risk falls within the scope of the notaries' insurance contract's general clause covering "the financial consequences of civil liability incurred... in the normal course of their duties as a result of their acts, faults or negligence, or the acts, faults or negligence of their staff". Currently, only atypical situations, such as a public official speculating on digital assets from client accounts or engaging in prohibited banking activities, would be covered by the legal and contractual insurability of criminally punishable activities. However, it should be noted that the French CSN (Conseil Supérieur du Notariat – High Council of Notaries) recently had the terms of its *ad hoc* insurance policy clarified by the French Ministry of the Interior's framework and planning law No. 2023-22 of 24 January 2023, which added a new article L. 12-10-1 to the French Insurance Code. This policy was taken out to protect the Central Register of Electronic Documents (Minutier Central des Actes Électroniques) against hacking, among other things. Despite some ambiguity in the wording,[1] it appears that this text confirms insurance coverage for ransomware attacks, provided that the victim files a complaint within 72 hours of the cyberattack. This applies not only to the CSN, but also to data-holding structures, including law firms, notary offices, and judicial officers. On a different note, the deterioration of a professional blockchain's security level for environmental reasons constitutes an increase in risk that may not always be declared, thereby exposing the insured party to the penalties set out in Article L. 113-4 of the Code des assurances. However, let us now return to the responsibility of legal professionals, although everything is linked.

2. The duty of legal professionals to provide advice within their digital environment

2.1. Do digital tools change the standard of a diligent and competent practitioner?

At first glance, and probably in principle, the answer is no. It is customary to state that the notary's duty of advice, later joined on this point by that of the lawyer, is absolute. The main consequence of this is that the practitioner's diligence cannot be weighed according to the client's competence or the level of personal assistance they provide. It should be emphasised that an 'augmented' client whose competence has been enhanced by seeking assistance from a legal tech company using generative AI – albeit on the fringes of the legal profession's advisory role and monopoly – would have the same rights to information, advice and investigation as a client without such assistance. Similarly, the assistance that a professional would receive when providing advice by using generative AI, or even quantum intelligence one day, is irrelevant to the level of diligence expected of them. This applies when a

[1] Article L. 12-10-1 of the French Insurance Code relates to "loss and damage" caused by an attack on a system. There is doubt over the inclusion of ransomware.

specialist is consulted or when the French Court of Cassation rules on the use of a notarial advisory body, such as a 'CRIDON'.[2] It should be noted that, while case law requires a strong convergence, if not assimilation, between civil and disciplinary sanctions, it is impossible to transpose ethical standards such as those contained in Article 22 of the French Code of Ethics for Notaries ('the notary owes his clients his professional conscience, consideration, impartiality, probity, advice appropriate to their situation and the most complete information') or Article 3 of Decree No. 2005-790 of 12 July 2005 ('shall demonstrate competence, dedication, diligence and prudence towards their clients') to the use of generative AI, regardless of its degree of depth and individualisation. A chatbot cannot determine what an 'impartial' opinion means, which prevents a notary from highlighting the economic inappropriateness of a deed they receive. Similarly, AI does not understand the fundamental concept of contractual loyalty. It is also impossible to make software understand the tact expected of a lawyer.

Since the standard of behaviour is not intrinsically altered by the digital environment, it is the increase in the amount of data available online and the performance of generative AI that will extrinsically influence the scope, if not the intensity, of the faults attributable to legal practitioners.

Moving away from the above technologies, the emergence of digital assets such as bitcoins and NFTs renews notaries' duty to inform and advise. They must also be aware of, and report, their extreme volatility in the event of donation sharing or a reportable donation, as this could affect the initial balance or the amount of debt to be reported. However, in order to provide effective advice, practitioners must be competent in decision support systems, given the many risks of misinterpretation that still exist, and perhaps always will, with generative AI and other processes supporting what remains, according to the current rules, a personalised intellectual service. Overall, the rapid progress of AI should not distract from the fact that, while it frees up practitioners to focus on their clients, the latter's expectations are in line with this qualitative improvement. They cannot ignore this improvement, not least when it comes to justifying the billing of services provided and liability. From this point of view, the qualitative gain is comparable to that brought about by long-established digital tools, such as public databases (e.g. Légifrance) and private databases (e.g. Lexisnexis, Lextenso, Lamyline etc.), which practitioners now rely on to keep up to date with the latest developments in positive law, rather than waiting for a journal to arrive in their letterbox. While a lack of mastery of IAG is likely to constitute professional misconduct on the part of a legal professional regardless of the degree of negligence, due to the principle of civil liability, it should be emphasised that the perfection of AI can also disrupt a profession by making the minimisation of liability risk almost irrelevant. This is not to mention the mind-boggling naivety of a Mr Schwartz[3] when faced with the unbridled creativity of

[2] *e.g.* 'Centre Régional d'Information et de Documentation Notariale'; Cass. 1st civ., 26 Oct. 2004, no. 03-16.358 for the Cridon opinion: "The personal skills or knowledge of the client, as well as information or opinions provided by third parties, cannot exempt the notary from their duty to provide advice, which is not relative in nature" – Cass. 1st civ., 27 Nov. 2019, no. 18-22.147: JurisData no. 2019-021437; P. PIERRE, in *JCP N*, 21-22, 2020, 1118: the presence of a tax advisor during donation-sharing is irrelevant.

[3] U. BECHINI, *AI, notaries and Maître Schwartz*, in *JCP N*, 2023, 1205: "AI could also help notaries to combat their biases. In my nearly 30 years of practice, I have never advised a client to set up a trust. This is, of course, a mistake that highlights my bias against trusts. Generally speaking, however, I do not regret this. Italy has no national law on trusts, so they are created under the law of another country. Clients must seek professional assistance with the relevant foreign law and, if necessary, take legal action under that law; the costs can be extremely high. Furthermore, the interaction between Italian and foreign law can have unpredictable consequences. Having said

ChatGPT. The failure of the French decree of 27 March 2020, which authorised the creation of a personal injury compensation algorithm called 'DataJust', was undoubtedly not only due to the difficulty of quantifying these complex damages. The cold, even hostile, reception from certain sections of the legal profession can also be explained by the fear of losing legitimacy in the face of the threat of other actors, such as victims' associations or even the victims themselves, gaining rapid, exhaustive and free nationwide knowledge of personal injury law.

In truth, a paradigm shift in professional liability law will only occur if the standard of a normally diligent and competent practitioner incorporates a new obligation to utilise all available digital tools. This shift from a right to a duty, and from an obligation of means to use digital technology to an obligation of result, could render legal professionals liable if they cannot prove that they have used the available digital tools. This trend can already be observed in other legal systems, such as in Canadian case law, which takes a very strict stance towards legal practitioners in this regard. This is not yet the case in France, where the Court of Cassation[4] requires professionals such as notaries to consult accessible legal notices, even if they are dematerialised. The Court censured a decision that criticised a public official for not consulting the Google search engine, stating that this would have enabled the official to question the seller's real situation by consulting documents relating to his company, which would have revealed the existence of liquidation proceedings. This would have enabled the official to question the seller's situation by consulting documents relating to their company, which would have revealed the existence of liquidation proceedings. However, the battle between paper and digital media is far from over. The Court of Cassation also ruled that a notary cannot escape liability by arguing that the government website they consulted had not been updated for several months about pollution risks when the correct information was available in a traditional register at the government's headquarters.

2.2. Are practitioners who use digital tools definitively liable?

Whether one considers liability or the corresponding insurance, the current state of the law does not place clients who are victims of poor digital practices by legal professionals in a precarious position. This is provided they can demonstrate that the information or advice provided was the result of the legal professional's insufficient knowledge, or their culpable tendency to place excessive trust in the machine. This tendency causes them to forget that the machine is a tool for dialogue, not monologue. However, these practitioners may be able to demonstrate the failure of the tools used by producing inaccurate data through imprecise algorithms or exaggerated cognitive biases, or due to the obsolescence of the legal data used or the technical impossibility of achieving the contractual objectives. Alternatively, bugs may compromise the normal functioning of the machines.[5] Such failures will not impact the rights of

---

that, perhaps in one or two cases, a trust would still have been the right solution. Had IA proposed a draft trust, I might have changed my mind".

[4] Cass. 1re civ., 28 nov. 2018, n° 17-31.144, JCP N 2018, n° 50, act. 938, obs. P. PIERRE.

[5] For example, see the note published on 12 February 2024 by the European Commission for the Efficiency of Justice (CEPEJ), which lists what legal professionals can expect from the use of IAG tools and their limitations (CEPEJ-GT-CYBERJUST (2023)5final). The note highlights several dangers, including the potential production of inaccurate information, the possible disclosure of sensitive data, potential violations of intellectual property and copyright, the limited ability to provide the same answer to an identical question, the potential reproduction of results, and the exaggeration of cognitive biases. The note concludes with a guide to the proper use of AI in a

victims if they can demonstrate their loss, which is often a missed opportunity for advice, and the causal link with the use of AIG or any other digital tool. The quality of information and advice is assessed independently of the process by which it is produced. It should be noted at this stage that notaries and lawyers cannot claim subsidiarity of their civil liability according to a traditional ruling of the Court of Cassation for the former, which has more recently been applied to the latter.

Therefore, the failure of a digital assistant can only be assessed once the client has been compensated. This will most often concern the professional liability insurer, who is subrogated to the client's rights. However, in the event of the practitioner not being insured, it will concern the collective notarial guarantee, or even the professional themselves if this second line of defence does not exist, as is the case for lawyers. In any case, these claims for recourse appear, at first glance, to have a solid foundation.

Firstly, this primarily concerns the liability of manufacturers for defective products, as set out in the European Directive of 23 October 2024.[6] Secondly, it concerns the proposal for a European directive on adapting rules on non-contractual civil liability to the field of AI.[7] Even though it does not directly address civil liability, the framework for AI provided by EU Regulation 2024/1689 (the 'AI Act') should also be considered, as it imposes obligations on providers according to the intensity of the risks involved. At first glance, the first directive appears to be very favourable to claims for recourse, since it now includes *"*digital manufacturing files" and "software*"* in defective products if they do not offer "the safety that a person can legitimately expect" (Art. 7/1 Dir.), and naturally because it continues to establish a strict liability regime against manufacturers in the broad sense.

However, the '*safety*' envisaged in this text is not consistent with the harm associated with the faults of legal professionals, whether the damage is bodily or material (Art. 6 Dir.).[8] Product defect does not appear to cover mere information failure, and the risk of product development remains a cause of exemption for its manufacturer.[9] What more can we expect from the proposed AI liability directive, which supplements the AI Act in this respect? It provides a fault-based liability regime for AI systems, whether high-risk or not, and gives the relevant authority the power to order the supplier or deploying entity to provide evidence, as well as establishing various presumptions of causality. However, for customers wishing to abandon primary proceedings against a practitioner and for practitioners acting on a recourse basis, such liability could be precluded *"*where the damage is caused by a human assessment followed by a human omission or act, and where the AI system has only provided information or advice

judicial context. B. DEFFAINS also mentions that ChatGPT confuses the 1993 Sapin 1 and 2016 Sapin 2 laws, as well as the Transparency Directive, with the MIF 1 and Prospectus Directives. It also claims that a real estate agent can purchase a property that they are responsible for selling. However, Article 1596 of the Civil Code prohibits agents from purchasing property entrusted to them for sale (B. DEFFAINS, *ChatGPT et le marché du droit*, in *JCP G,* 2023, doctr. 430).

[6] EU Directive 2024/2853, 23 October 2024.

[7] Doc. COM 2022, 496 final, 28 Sept. 2022. This proposal has been withdrawn from the European Union's work programme for the time being. It is unclear whether new regulations will be proposed in the future.

[8] This does not apply to the loss or corruption of data that is not used for professional purposes.

[9] M. JULIENNE, *Les directives intéressant la responsabilité en matière d'IA*, in *Rev. dr. banc. et fin.,* 2024.

that has been taken into consideration by the relevant human actor" (Recital 15).[10] This is precisely the case with the exemption from advice by a legal professional when using AI, unless the machine replaces the latter. This will not apply to advice, but it could apply to the technical diligence expected of legal professionals in the long term. In any event, as with ChatGPT, it seems that the proposed directive is unable to neutralise the general terms and conditions of most AIG systems, which currently exclude any liability for incorrect or false information.[11]

3. Technical diligence of legal professionals in their digital environment

3.1. Does distancing documents alter liability terms?

The role of legal professionals is being called into question by the development of remote document signing by lawyers and notaries. French law No. 200-230 of 13 March 2000 made it possible to sign such documents electronically, establishing their probative value as being equivalent to that of paper documents (Civil Code, Art. 1366). This law is now widely accepted by legal practitioners.

With regard to the assessment of the liability of public officials,[12] a document in electronic form is subject to the same formal requirements and penalties for irregularity as its paper counterpart.[13] The parties' signatures are merely reflections of their handwritten signatures, entered via a digital tablet, on the 'visible on screen' document. Remote appearance for powers of attorney and certification by Docusign does not prevent each professional involved from personally verifying the signatories' identities and consent. The electronic signature of the public official, affixed by means of an encrypted and secure Réal key, does not change in nature or function. It certifies the regularity of the document's preparation with a view to establishing its authenticity (Civil Code, Art. 1367). In other words, *"*it is the instrument that changes, not the person using it or the diligence that can be expected of them when acting as a public official".[14]

However, a new stage in the dematerialisation of authentic documents, emerged from health requirements : the remote appearance of the parties to the document was promoted temporarily by the decree of 3 April 2020[15] and was established definitively, by the decree of 20 November on powers of attorney.[16] Article 1 of the former decree stated that "the exchange of information necessary for the

[10] In addition, in the latter case, it is possible to link the damage to human error because the AI system did not affect the outcome, and establishing causality is no more difficult in such situations than in situations where no AI system was involved. A return to ordinary liability is therefore necessary.

[11] B. DEFFAINS, *ChatGPT et le marché du droit*, cit., 37.

[12] For lawyers, see S. AMRANI-MEKKI, M. MEKKI, *Avocat et intelligence artificielle: l'intelligence artificielle propose, l'avocat dispose*, in *Rev. prat. prospective et innovation*, 1, 2024. They note that "many software programmes equipped with artificial intelligence can improve secure remote identification procedures (such as Remote ID, as seen in ID360) and collect relevant information through document analysis (e.g. Autolex and Lexis+AI)".

[13] See Acts of the 111th Congress of Notaries of France, *Legal certainty*, Strasbourg 2015, 169 ff. for details on Decree No. 2005-973 of 10 August 2005, which amends that of 26 November 1971 on acts drawn up by notaries, governing both paper and electronic documents.

[14] L. AYNES, *Authenticity*, in *La documentation française*, 124, 2013, 158.

[15] French Decree No. 2020-395 of 3 April 2020 authorises remote notarial acts during the health emergency.

[16] French Decree No. 2020-1422, introduced on 20 November 2020, concerns remote notarised powers of attorney.

drafting of the document and the collection, by the notary acting as instrument, of the consent or declaration of each party or person involved in the document, shall be carried out by means of a communication and information transmission system guaranteeing the identification of the parties, the integrity and confidentiality of the content, and approved by the High Council of Notaries". This digital distancing – both physical and chronological – disrupts the identification method of the parties, which is now delegated to an external certification authority. This authority uses a combination of email and SMS exchanges to secure the link with the document to be signed. Furthermore, the collection of consent and the signature expressing it are based on an electronic certificate sent to the notary by the same certifying third party. The public official then digitally signs the document in accordance with the procedure familiar to practitioners since the 13 March 2000 law no. 2000-230. The implications of this leap forward[17] in terms of professional liability are striking. In fact, the French Court of Cassation rigorously monitors the direct collection of signatures, identity checks and, more generally, any "apparent anomaly".[18] Case law regularly addresses the fundamental role of the public official in verifying the capacity and consent of the parties to the document through the prism of civil liability, a role which is inevitably complicated by distance.[19] In an age of neurodegenerative diseases, it will not always be possible to assess the real or apparent lucidity of a client through an immaterial meeting. Regardless of their receptiveness, it is difficult to provide clear, precise and appropriate explanations to a client, which professionals sometimes perceive as being more intuitive than rational. While the idea of a third party lurking in the shadows of a remote signatory, compelling them to comply, is undoubtedly a contractual fantasy, the same cannot be said for a state of psychological or even economic dependence[20] which is sometimes perceptible through subtle symptoms. Along these lines, courts often consider the evidence available to the notary when assessing the veracity of a client's statements. This duty of suspicion may be imposed on the public official based on an interview, which is even more effective when conducted without digital intermediation. In connection with the above developments (I), it should be recalled that the duty of advice of a legal professional is absolute and cannot be reduced by physical distance, and that liability is assessed in the abstract according to a standard that digital tools cannot change. From a comparative law perspective, these factors, which are likely to reduce the separation of acts, have probably caught the attention of Quebec lawyers. After widely opening access

[17] The French Conseil d'État did not grant the request to suspend the 3 April 2020 decree, ruling that 'there is no legislative provision stipulating that notaries can only perform their duties in the presence of the parties, and the decree merely temporarily derogates from the procedures set out in the 26 November 1971 decree on deeds drawn up by notaries, under which public officials may draw up authentic deeds': CE, 15 April 2020, No. 439992, Defrénois 23 April 2020, No. 17, 10.

[18] Cass. 1st civ., 29 May 2013, no. 12-21.781, M. MEKKI, in *JCP N*, 49, 2013, 1282, no. 19: the notary was criticised for failing to suspect that the transfer orders, which were apparently issued by the mortgage borrower, were false. He could have compared the signature on these orders with that on the loan agreement drawn up in his office. Cass. 1st civ., 2 Oct. 2013, no. 12-24.754, P. PIERRE, in *JCP N*, 2014, 1194. Also see the ruling on the absence of any apparent anomaly in the minutes of a general meeting, which rules out the need for signature certification: Cass. 1ère civ. 26 Feb. 2020 no. 18-25.671.

[19] M. MEKKI, *L'intelligence artificielle et la profession notariale*, in *JCP N*, 2019, 1001. He rightly points out that "the issue is all the more sensitive given that there is now a new form of illiteracy: digital illiteracy".

[20] Article 1143 of the French Civil Code.

to remote notarial appearances, a law on 24 October 2023 reversed this principle, making it a mere exception.[21]

3.2. Could the legal professionals' liability protect them against the unbridled digitisation of their work?

The paradox is as follows: the ever-increasing performance of the various auxiliary systems available to legal professionals could ultimately compromise their legitimacy by turning them into direct competitors and undermining the safeguards of the legal system. Nevertheless, civil liability law may ultimately prove to be the best protection for practitioners against this insidious competition.
On the lawyers' side, tools such as the *'e-co-pilot'*, which helps them *"*conduct research and due diligence using natural language instructions, leveraging the OpenAI model', automate 'various aspects of legal work, such as contract analysis, due diligence, litigation and regulatory compliance', and 'automatically generate procedural documents (such as summonses)".[22] Similarly, it has been suggested that

> the integrity of contractual content is reinforced by digital tools in the field of contract analytics (Legisway, Henchman, Dilitrust, etc.). On the one hand, these tools can perform normativity checks using jurimetrics to alert users when a clause is no longer compliant with the current legislation, is ineffective in a system whose effectiveness is called into question by controversial case law, or is threatened by a national or European reform project currently under discussion. This normativity check is coupled with a compatibility check. In this case, the AI system can compare clauses in the same contract to detect any inconsistencies (Civil Code, Art. 1119), or clauses in several contracts if the contractual relationship is long-term. It can also compare their respective general and specific terms and conditions.[23]

Such exponential growth in performance could call into question significant areas of lawyers' specific expertise. It is no longer a question of distancing themselves from practitioners, but of replacing them.
According to their promoters, the automation of drafting processes and their extensions could reduce the risk of human error and the associated liability for notaries. This is where the potential of blockchain technology and its various applications, such as smart contracts, comes in. In short, the aim is to store and secure data, authenticate exchanges, and guarantee their infalsifiability and indestructibility.[24] In fact, cross-referencing data using these digital blockchains provides significant resources for legalisation and certification, ensuring the integrity of documents, enabling traceability of associated digital

[21] The law is aimed at modernizing the notarial profession and promoting access to justice. (Bill 34 or Law 23, which came into force on 24 October 2023). N. CHAPUIS, *Recul de la signature à distance des actes authentiques au Québec, contexte et constats*, in *JCP N*, 2024, 1084, Emphasizing that the notary may use remote signing under the following cumulative conditions: "If the client requests it; if circumstances require it; and if the signature can be made in accordance with the rights and interests of the parties. This exceptional procedure must not become customary".

[22] B. DEFFAINS, *ChatGPT et le marché du droit,* cit., 17, also citing Lawdroid Copilot, a tool for drafting legal documents that uses GPT-3 to generate contracts and confidentiality agreements based on information provided by the user

[23] S. AMRANI-MEKKI, M. MEKKI, *Avocat et IA*, cit., 9.

[24] C. ROSSIGNOL, *Blockchain: quel futur pour les notaires?*, in *Le Club du droit*.

fingerprints and preserving them. Thanks to the security offered by structurally tamper-proof consortium blockchains, avenues that can be explored include the issuance of copies of authentic documents and notarised certificates by public officials, as well as the inclusion of technical documents required for the sale of real estate in annexes, or even written or filmed explanations of last wills and testaments. This touches on the very essence of the notary's mission, as the development of blockchains and smart contracts could undermine the legitimacy of public officials[25] by fulfilling two of their essential functions, preservation and certification.

Nevertheless, the operation of consortium blockchains remains dependent on the actions of individual practitioners. The absolute traceability of transactions in the form of digital fingerprints could greatly simplify the process of proving the diligence performed by these practitioners, whether technical or informational. Conversely, if it transpires that the data has been entered incorrectly or insufficiently, or that the digital resource has remained unexploited, liability proceedings against professionals will be facilitated. The same observation applies to lawyers, and it has rightly been pointed out that "the drafting lawyer retains a decisive role both in and out". Upstream, the prompting lawyer must use their experience and expertise to decode the parties' expectations and the strategy to be pursued. They will then formulate these in relevant prompts. Outside of this process, once the work has been produced by the machine, the supervising lawyer must analyse the result, question it, correct it, supplement it, or even simplify it.[26]

However, the dynamics of blockchains are pushing them towards public blockchains, which are freely accessible. The trend towards automating procedures goes hand in hand with depersonalisation. At first glance, the advantages of consortium blockchains seem to extend to the public, who would avoid the cost of intermediation by a notary or other legal practitioner while enjoying the security and reliability of digital blockchains. For example, certification of electronic signatures by a trusted third party[27] could be the first step towards disintermediation through digital fingerprinting by blockchains. Complete disintermediation is not unheard of; it has already been explored by African countries without land registries and even by European countries. For example, Sweden is considering such a change for the registration of property transfers.[28] That said, this prospect is only possible because there are no public officials responsible for prior checks, and because the Swedish state covers any damages resulting from deliberate or unintentional inaccuracies directly. Clearly, the development of blockchains is inseparable from the risk management model involved in verifying their content, which lies at the heart of the notarial authentication process.

Absolute disintermediation would therefore imply the complete removal of responsibilities,[29] which is difficult to conceive in the absence of solvency guarantees currently offered by individual civil liability insurance being transferred to a public structure. Furthermore, as previously explained, the backup

[25] B. BEIGNIER, A. TANI, *Le notaire et le testament olographe*, in *Dr. famille*, 2018, 12: "By offering the same services as online sites, notarial services risk being unable to stand out and may eventually suffer from the 'Uberisation' of legal tech in sectors that do not fall within their exclusive competence. We remember the hesitations when the testamento.fr site was launched".

[26] S. AMRANI-MEKKI, M. MEKKI, *Avocat et intelligence artificielle*, cit., 10.

[27] *Supra*, I/B.

[28] T. VACHON, *Notariat du 21ème siècle, enfin le zéro papier?* 48ème Congrès du MJN, note 42, 243.

[29] M. MEKKI, *L'intelligence artificielle et la profession notariale*, cit., 55 and 56.

guarantees that could be provided by insurers of digital tool designers pale in comparison to those of their users as they are very difficult to mobilise given the current state of liability law.

4. Conclusion

In conclusion, we wholeheartedly support the significant decision made by the CJEU, which emphasised that individual responsibility is a key factor in determining the legitimacy of legal practitioners.[30] This responsibility is linked to a professional status that cannot be reduced to anything else, regardless of whether practitioners use digital tools.

[30] CJEU, 24 May 2011, Commission v France, C-50/08.