An examination of fundamental rights protection in AI governance

Laura Evangelista
INAPP

Concetta Fonzo
INAPP

Eleonora Zecca
LUISS

This article provides an analysis of the impact of Artificial intelligence (AI) on contemporary society, highlighting the numerous violations of fundamental rights, including the rights to privacy, non-discrimination, and access to information. Drawing on concrete examples and in light of the inadequacy of current legislation, the paper examines existing jurisprudence on AI, beginning with the *Artificial intelligence Act* adopted by the European Commission in 2024 and comparing it with U.S. legislation to evaluate differing regulatory approaches. The article concludes by discussing the future challenges involved in developing a trustworthy and reliable AI system capable of safeguarding fundamental rights and ensuring algorithmic accountability.

L'articolo offre un'analisi dell'impatto che l'Intelligenza artificiale (IA) ha sulla società odierna, nonché delle numerose violazioni di diritti fondamentali, come il diritto alla privacy, alla non discriminazione e all'informazione. Alla luce di esempi concreti e della legislazione attuale insufficiente, il contributo esamina la giurisprudenza esistente sull'IA, a partire dall'Artificial intelligence Act adottato dalla Commissione europea nel 2024, confrontandolo con la legislazione degli Stati Uniti per valutare i diversi approcci. Il lavoro termina con una disamina delle sfide future verso un sistema di IA affidabile, che garantisca la protezione dei diritti e algoritmi attendibili.

DOI: 10.53223/Sinappsi_2025-02-6

Citazione
Evangelista L., Fonzo C., Zecca E. (2025), An examination of fundamental rights protection in AI governance, *Sinappsi,* XV, n.2, pp.75-89

Parole chiave
Artificial intelligence
European policies
Fundamental rights

Keywords
Intelligenza artificiale
Politiche comunitarie
Diritti fondamentali

Introduction

In just a few decades, Artificial intelligence (AI) has shifted from the realm of science fiction to a significant force influencing nearly every facet of contemporary life, redefining the limits of machine capabilities. Central to this technological transformation is a paradox: AI possesses the potential to significantly improve human lives in ways never seen before, yet it also introduces substantial risks related to privacy, misinformation, and discrimination. To date, scholars refer to AI as the development of computer systems designed to perform tasks that typically require human intelligence, such as visual perception, speech recognition and decision-making[1], although a

1 On this point, for instance, see: Copeland B.J., Artificial intelligence, in *Encyclopedia Britannica,* available at: https://www.britannica.com/technology/artificial-intelligence.

universally accepted definition of artificial intelligence (West 2018) still seems to be lacking.

Moreover, while many might assume that studies on AI began recently, Alan Turing is often recognised as the pioneer of the concept. In 1950, he theorised about "thinking machines" capable of reasoning like humans. His "Turing Test" stated that computers must complete reasoning tasks at a level comparable to humans in order to be considered "thinking" in an independent manner (Turing 1950). Later, building on Turing's ideas, McCarthy was the first to introduce the term "Artificial intelligence" to describe the field focused on creating intelligent machines, particularly software. According to McCarthy, this field is linked to the challenge of employing computers to understand human intelligence (McCarthy 2022). The issue, however, is that it's still impossible to define what types of computational processes can be classified as intelligent, as the mechanisms underlying human intelligence are not yet fully understood (*ibidem*). These challenges became apparent with the development of the first AI machines. The early wave of Artificial intelligence proposed a "Symbolic AI" system which aimed to create computers that could mimic the problem-solving skills of the human brain[2]. A key application of "Symbolic AI" was developed at the Massachusetts Institute of Technology in 1966 by Joseph Weizembaum, who developed *ELIZA,* a chatbot programme designed to simulate conversations between a patient and a psychotherapist (Treusch 2020). *ELIZA* was based on a script containing rules, and input sentences were restricted to simple declarative and interrogative forms based on known objects and relationships. In order to communicate, *ELIZA* used patterns and predefined rules to generate responses (*ibidem*). Despite general enthusiasm for Artificial intelligence, the "Symbolic AI" approach ultimately failed. It became clear that many fields or domains — language, for instance — cannot be precisely formalised through rules and symbolic manipulation. This first attempt at AI faced significant criticism. Hubert Dreyfus (1972) argued that "Symbolic AI" failed because it was based on philosophical assumptions that compared the brain to a computer (Kenaw 2008). He argued that the brain does not operate like a computer executing distinct tasks, nor does the mind process information solely based on formal principles. Dreyfus also questioned the possibility that all knowledge could be completely articulated in logical or mathematical formats and that reality could be broken down into discrete facts represented by symbols (*ibidem*).

After an "AI winter", a renewed effort to advance the field emerged with the "Connectionist AI", which theorised that machines should simulate neural networks[3] (Hardesty 2017), inspired by the brain structures and functions, rather than relying on human reasoning and logic, as "Symbolic AI" had unsuccessfully attempted (Smolensky 1987). This new approach was more promising, as it was data-driven and capable of learning from examples rather than being rule-based. Following this approach, Frank Rosenblatt developed *The Perceptron* (1958), one of the earliest artificial neural networks. It used a system of interconnected units (like biological neurons) to adaptively classify inputs (Block 1962). This early application of "Connectionist AI" aimed to represent intelligence through networks of simple processing units. In contrast to "Symbolic AI", which employed explicit rules, "the Perceptron" relied on probabilistic models, allowing it to learn and adjust by forming new associations among its components (*ibidem*). However, the main limitation of Perceptron's single-layer structure was in identifying only linearly separable patterns, a limitation later overcome by the development of multi-layer neural networks. Geoffrey Hinton, widely regarded as the "Godfather of Deep Learning", played a crucial role in advancing neural network research. He was able to develop unsupervised pre-training for deep networks, which proved instrumental in maintaining progress in the field and laid the foundation for modern deep learning (LeCun *et al.* 2015). Through his works, Hinton could demonstrate the power of multi-layer

2 For an overview of symbolic AI's evolution, from its philosophical foundations in the 1950s to its modern-day fusion with neural networks, see: *The Evolution of Symbolic AI: From Early Concepts to Modern Applications*, available at: https://smythos.com/developers/agent-development/history-of-symbolic-ai/.

3 Definition: "neural nets are a means of doing machine learning, in which a computer learns to perform some tasks by analysing training examples. Usually, the examples have been hand-labelled in advance". Retrieved from: https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414.

neural networks[4] (deep neural networks), and their ability to learn complex, non-linear patterns and derive hierarchical features directly from raw data. In recognition of his contributions, he was awarded the Nobel Prize in 2024.

Over the years, AI has made significant progress, achieving the appearance of intelligent behaviour by operating through algorithms[5] and statistical learning techniques on extensive datasets (Neapolitan 2012). This made possible concrete advancements, such as the development of "Large Language Models (LLMs)", which generate text that mimics human writing and supports various tasks (Blank 2023).

AI rapid advancement, however, poses significant challenges to fundamental rights. AI-driven systems increasingly threaten privacy and contribute to misinformation and discrimination. Large Language Models often amplify biases already present in their data and propagate harmful stereotypes and hegemonic narratives, despite lacking true understanding of language or possessing real-world knowledge, leading to potential misuse. This occurs because LLMs are pre-trained using massive amounts of data coming from publicly available internet sources, thus often providing low-quality and unrepresentative data, contributing to the spread of misinformation. This has prompted concerns about potential harmful data colonialism by big AI companies, which can easily extract human life data, tracking consumers behaviours, targeting them with personal-branded online content (Couldry and Mejias 2019). This form of mass surveillance contributes to spreading misinformation through AI-generated content, undermining public trust and democratic processes, reinforcing discrimination, amplifying biases and inequalities.

Starting from such a scenario, this article aims to reflect on the impact of artificial intelligence on society, focusing on how its adoption may harm fundamental rights, such as the right to privacy, fair access to information, and non-discrimination, providing concrete case studies as application of such violations in the use of AI tools. This contribution will explore and examine how the newly adopted European Artificial Intelligent Act (European Union 2024, hereinafter AI Act) seeks to protect individuals from the disproportionate use of AI tools, comparing it with legislative tools adopted in the United States, in order to evaluate different legislative approaches and measures. Finally, this article will outline future challenges and goals for improving the development of Artificial intelligence, aiming for a trustworthy AI system that ensures rights protection and reliable safe algorithmic functions. Nevertheless, this challenge is still at the very beginning. At present, it is not possible to develop robust AI systems capable of operating reliably under all conditions, therefore human oversight remains essential.

1. Artificial intelligence and rights: opportunities and threats

Even though AI technologies hold the promise of improving the protection of human rights and the rule of law, they also pose significant risks, potentially leading to discrimination, gender inequality, threats to democratic process, and violations of human rights (Council of Europe 2024). Firmly aware of these threats, the European Union has been committed to ensuring that the development of AI aligns with established human rights principles. In line with this commitment, the Steering Committee for Human Rights (CDDH) has been tasked by the Committee of Ministers with focusing on this emerging area, establishing the Drafting Group on Human Rights and Artificial intelligence (CDDH-IA), which is responsible for creating a Handbook on Human Rights and Artificial intelligence by the end of 2025[6].

Examining the influence of AI on human rights is complex. Many documents primarily focus on identifying the rights and freedoms that could be affected, rather than actively evaluating this potential impact and suggesting assessment frameworks (Mantelero 2022). For this reason, the

4 Definition: "a multilayer neural network has multiple layers stacked on top of each other. Each layer receives input from the previous layer and applies a mathematical operation called an activation function, such as the sigmoid function. This allows the network to capture complex relationships between inputs and outputs". Retrieved from: ttps://muneebsa.medium.com/deep-learning-101-lesson-9-multi-layer-neural-network-7cd3a53066c8.

5 Definition: "a set of mathematical instructions or rules that, especially if given to a computer, will help to calculate an answer to a problem". Retrieved from: https://dictionary.cambridge.org/dictionary/english/algorithm.

6 On this point, see: Council of Europe, *Artificial intelligence and Human Rights (CDDH-IA)*, available at: https://www.coe.int/en/web/human-rights-intergovernmental-cooperation/intelligence-artificielle.

present article analyses the risks AI systems pose to human rights from a legal perspective, setting aside ethical concerns, while providing a case study for each type of rights violation to better understand real potential implications and consequences.

Artificial intelligence and the right to information

The rise of new Generative AI (GAI) technologies[7] can manipulate and create several types of digital content following specific guidelines easily enabling multi-modal misinformation (Rombach 2022). For instance, the increasing amount of Artificial intelligence Generated Content (AIGC) hinders the battle against misinformation (Sohail *et al.* 2023). However, within the realm of fake news, it is necessary to differentiate between disinformation and misinformation. Disinformation is defined as "false, inaccurate or misleading information" spread to deceive the recipients, whereas misinformation "refers to false, inaccurate or misleading information that is shared without any intent to deceive" (Bontridder and Poullet 2021). When AI is involved in the creation of false or fake content, the result is known as a "deepfake"[8]. More specifically, deepfakes are the result of two AI algorithms operating in a so-called Generative Adversarial Network (GAN), a method for algorithmically generating new data from existing datasets (*ibidem*). AI-driven personalisation algorithms have emerged as powerful yet problematic amplifiers of misinformation across social media platforms. This is due to the economic model adopted by the major platforms, known as "the economics of attention" (Festré 2015), whereby investments are indirectly financed through revenues from advertising firms (Bontridder and Poullet 2021). Consequently, the more time users spend on a platform, the more that platform profits (*ibidem*). Users, by interacting with the platform, inform the algorithm about their preferences, what they like, disagree with, think or feel. As a result, the algorithms recommend content that closely aligns with the users interests and opinions, which drives greater profits but, at the same time, arms the user, who ends up seeing and reading only one perspective. In addition, the echo chambers and filter bubbles play a pivotal role in the spread of misinformation and disinformation. The former, often created as a consequence of the latter, refer to contexts in which people are exposed exclusively to viewpoints that reinforce their own beliefs, disregarding opposing perspectives (Festré 2015). This is exactly what occurs with AI-driven personalisation algorithms. Another factor to consider in this discussion is Generative AI, such as ChatGPT and Stable Diffusion, which are increasingly popular, sophisticated algorithms capable of producing highly realistic and convincing content across various forms, including text, images, audio, and video (Seow *et al.* 2022). A key example of Generative AI application has been the spread of fake robocalls from Joe Biden in New Hampshire (Hsu 2024). Using Generative AI, a company had published fake recordings of US President Joe Biden's voice, suggesting New Hampshire voters should not vote (Associated Press 2024). A different US-related episode occurred during the 2016 US Presidential elections, when algorithms, automation, and GAI were used to increase the effectiveness and reach of disinformation campaigns and related cyber activities, which influenced the opinions and voting decisions of American citizens (Kertysova 2018). The combination of multiple modalities by generative models poses a challenge in the fight against false information, making it more difficult for both users and algorithms to distinguish between reality and deception (Xu *et al.* 2023). For this reason, countermeasures have been adopted at the user, platform and government-level. User-level resources include educational and verification websites, while social media platforms such as Facebook, Twitter, and YouTube actively label, delete, limit, pre-emptively counter, and provide information regarding misinformation (Waldemarsson 2020). The European Union's approach against disinformation seeks to find a balance between conflicting constitutional rights and freedoms that converge in the sphere of disinformation. This approach has led to the adoption of the Digital Services Act, which strictly

7 Definition: "Generative AI can be thought of as a machine-learning model that is trained to create new data, rather than making a prediction about a specific dataset. A generative AI system is one that learns to generate more objects that look like the data it was trained on". Retrieved from: https://news.mit.edu/2023/explained-generative-ai-1109.

8 Definition (Article 3, AI Act): "means AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthful". Retrieved from: https://artificialintelligenceact.eu/article/3/.

regulates platforms involved in the dissemination of disinformation (Turillazzi *et al.* 2023). Additionally, the introduction of further regulatory measures aimed at collaborating with platforms to combat specific online content, such as the Enhanced Code of Practice on Disinformation, has played a significant role in shaping the European strategy against disinformation (Pamment 2020). This could be the first rigorous approach in tackling biases, and thus fake news, in defence of fundamental human rights against AI-driven threats. Bias is more related to discrimination, which violates human rights, as the following section will examine in greater depth.

Artificial intelligence and the right to non-discrimination

Generative AI, as previously mentioned, operates through algorithms, which are essential for streamlining numerous tasks that impact our daily lives, often on a scale beyond human capability. This is possible because algorithms work following computer instructions to convert inputs into outputs. Despite these 'superpowers' the potential for bias in algorithms is extremely high (Bozdag 2013). Even if machines and technologies appear impartial and neutral, they are not. In fact, people who develop algorithms may embody certain biases, which are easily transferable to machines. Therefore, the tendency of algorithms to generate discriminatory outcomes is not intrinsic to AI technology itself, but rather deeply rooted in societal factors influencing those who train these technologies (Nelson 2019).

Even if it seems that discrimination derives from a biased algorithm, it is necessary to distinguish between bias and discrimination. The term "bias" may have several meanings depending on the context. For the purposes of this article, it can be defined as: "unequal treatment based on protected characteristics, which includes discrimination and hate-motivated crimes" (Nelson 2019). Discrimination, on the other hand, refers to a situation where a person's disadvantage is based on one or more of these protected attributes (Nelson 2019).

Clearly, algorithmic bias results in discrimination when reliance on a protected characteristic causes a less favourable outcome, especially when coded parameters, training data, or input data feature elements that directly reflect a protected characteristic (European Union Agency for Fundamental Rights 2022). However, algorithmic bias most often leads to implicit discrimination, which is harder to identify and resolve.

Despite the evident risks that AI technologies pose, the use of algorithmic decision-making has been adopted in the U.S. criminal justice system to estimate potential recidivism. A prominent example of such a tool is the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) algorithm (Deb 2023). In 2016, the investigative journalism ProPublica conducted a thorough analysis of COMPAS, discovering racial biases in its risk assessments procedure. The original aim of COMPAS was to assist criminal justice professionals to assess possible future offenses (Brackey 2019). Even though COMPAS does not consider race as a factor to be calculated, in ProPublica's findings revealed disproportionate negative effects on Black defendants (Deb 2023). ProPublica examined over 7,000 cases between 2013 and 2014 to compare predicted recidivism scores with actual reoffending rates (Angwin *et al.* 2016). What emerged was deeply concerning: the algorithm's ability in predicting recidivism appeared to be very poor, only 20% of those labelled as high-risk actually reoffended (Angwin *et al.* 2016). The problem of discrimination emerged when the algorithm proved to be inaccurate for both black and white defendants. Black defendants were twice as likely to be wrongly classified as high-risk, whereas white defendants were more likely to be incorrectly classified as low-risk (Angwin *et al.* 2016).

Despite these alarming flaws, COMPAS was used in court sentencing decisions (Deb 2023). Afterwards, the founder of COMPAS acknowledged that the algorithm had not been designed for sentencing purposes, stating that: "the risk score alone should not determine the sentence of an offender" (Supreme Court 2016). Nonetheless, the indiscriminate use of this score in court trials raises profound ethical issues concerning automation bias, especially when human decision-makers rely exclusively on algorithmic results without critical evaluation.

To prevent such episodes, EU institutions have become increasingly engaged in addressing AI bias and other fundamental rights concerns. The AI Act (European Union 2024), a pioneering legislative proposal by the European Commission, regulates AI systems based on a risk-based framework. This

approach classifies AI applications into different categories, with the strictest regulations applied to high-risk systems, such as those used in employment, education, healthcare, and law enforcement. These systems must adhere to stringent requirements, including bias monitoring, transparency, and human oversight. The AI Act also prohibits certain AI practices outright, such as social scoring and manipulative techniques that could unfairly exploit individuals (*ibidem*, Article 5). Through these regulations, the EU aims to ensure that AI systems operate in an ethical, fair, and non-discriminatory manner. From another perspective, the General Data Protection Regulation (GDPR), which came into effect in 2018, plays a crucial role in shaping the ethical use of AI, particularly in the realm of automated decision-making (European Parliament and Council of the European Union 2016). Article 22 of the GDPR grants individuals the right to contest and opt out of fully automated decisions that have significant consequences on their lives, such as hiring processes or financial approvals (European Parliament and Council of the European Union 2016). Furthermore, the regulation enforces strict data protection principles, ensuring that personal data used in AI models is processed fairly and does not lead to discriminatory outcomes. Transparency is also a fundamental requirement under GDPR, as individuals have the right to be informed about how AI-driven decisions affecting them are made. The Digital Services Act (DSA)[9] and Digital Markets Act[10] (DMA) complement these efforts by targeting large digital platforms and ensuring that algorithmic decision-making processes remain transparent and accountable. These regulations aim to prevent discriminatory practices by compelling major technology companies to disclose how their AI systems operate, particularly in areas like content moderation, online advertising, and recommendation algorithms. By establishing these legal safeguards, the EU seeks to create a digital environment where AI-driven services respect fundamental rights and promote fairness.

Artificial intelligence and the right to privacy

In today's digital age the enormous quantities of data created and exchanged online allow businesses, governments, and organisations to improve their decision-making processes. Most of the time, this data includes sensitive information that individuals may prefer to keep private. This is where the concept of privacy comes into play. Privacy refers to the right to keep personal information secure and protected from unauthorised access (De Capitani Di Vimercati *et al.* 2012). This right has become increasingly significant due to the increasing volume of personal data being collected and analysed (DeVries 2003).

AI poses a risk to the privacy of both individuals and organizations, mainly due to the complexity of the algorithms used in AI systems and the fact that personal data is often a key component of the datasets used to train machine learning models. Additionally, AI technology is capable of making decisions based on subtle data patterns that often go unnoticed by humans. One of the most common issues consists in surveillance practices undermining personal autonomy and existing power disparities. This practice stems from the need to gather massive amounts of data to train AI systems, thus increasing the volume of personal data in AI systems and raising serious concerns about privacy and data protection.

Governments must therefore proactively safeguards individuals' privacy rights by implementing powerful data security measures, ensuring that data is used exclusively for its intended purposes. Furthermore, transparency in the use of personal data by AI systems is vital in order for users to understand how their data is being utilised.

Within the European Union framework, the adoption of the AI Act, specifically Article 78, imposes the confidentiality "of information and data obtained in carrying out their tasks and activities" (European Union 2024). The scope of this provision is to protect intellectual property rights, confidential business information, and trade secrets of both individuals and legal entities, including source code, as well as public and national security interests and the conduct of criminal or administrative proceedings. At another level, the General Data Protection Regulation (GDPR) outlines specific obligations for businesses

9 An official overview of the Act is provided by the European Commission at: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en.

10 An official overview of the Act is provided by the European Commission at: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-markets-act-ensuring-fair-and-open-digital-markets_en.

and organisations regarding the collection, storage, and management of personal data. It applies to both European organisations handling the personal data of individuals within the EU and to organisations based outside the EU that target individuals residing in the EU. Even though AI is not explicitly mentioned in the GDPR, AI systems must strictly comply with its requirements, especially when processing the personal data of EU citizens.

Numerous violations have occurred despite rigorous regulations designed to protect users. In 2022, Clearview AI, a facial recognition firm, became embroiled in one of the most significant privacy controversies in recent years. This company gathered billions of images from public websites without users consent to train its facial recognition technologies, enabling law enforcement agencies to match faces with identities into a database (Dul 2022). These practices raised serious concerns about violations of privacy rights, so much so that the Italian data protection authority launched an investigation following media reports highlighting multiple issues related to Clearview AÌs facial recognition products. In 2021, the Authority had also received further complaints from privacy and fundamental rights organisations expressing concern about Clearview AÌs activities. The investigative results highlighted several GDPR violations, including breaches of transparency, purpose limitation, and storage limitation principles (Pisanu 2022). Ultimately, the Italian data protection authority imposed a fine of EUR 20 million and the prohibition for any further web scraping activities within Italian territory. In addition, Clearview AI was required to appoint a representative within the European Union[11].

2. Introductory legal framework to Artificial intelligence

Until recently, authorities lacked the powers, procedural frameworks, and resources needed to ensure and monitor compliance with AI-related regulations, leaving national bodies solely responsible for enforcing safety and fundamental rights standards. This legal uncertainty and complexity discouraged businesses from developing and using AI technologies, slowing AI advancement in Europe and reducing the EUs global competitiveness in this sector. As already discussed, AI systems pose significant threats to fundamental rights such as non-discrimination, freedom of expression, human dignity, and the protection of personal data and privacy. The general framework of right protection under the EU Charter of Fundamental Rights and the General Data Protection Regulation (GDPR), was not sufficient to guarantee adequate safeguards against AI technologies (European Commission 2021). As a result, EU policymakers have committed to adopting a "human-centric" approach to AI, ensuring that Europeans can benefit from new technologies while upholding the EUs principles and values (Amariles and Baquero 2023). In its 2020 White Paper on Artificial intelligence (European Commission 2020), the European Commission pledged to promote the adoption of AI while also addressing the risks associated with certain applications of this emerging technology. Following an initial soft-law strategy, which included the 2019 Ethics Guidelines for Trustworthy AI and related policy and investment recommendations, the European Commission transitioned to a legislative approach (Madiega 2021). This development established standardised regulations for the development, market introduction, and application of AI systems (*ibidem*). The real breakthrough in AI legislation came in 2024 with the adoption of the AI Act, a targeted legislation capable of providing a clear legal framework to guide future AI development, ensuring compliance with ethical and legal standards while fostering innovation in a responsible manner.

Artificial intelligence act: the AI Act

In June 2024, European Union lawmakers signed the AI Act, establishing the first binding horizontal regulation on AI worldwide. The primary objective

11 A concise recap of the Clearview case is provided in a press release by the Italian Data Protection Authority, available at: https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9751323.
For the full decision issued by the Italian Authority against Clearview, see: Garante per la Protezione dei Dati Personali – GPDP (2022) *Ordinanza ingiunzione nei confronti di Clearview AI* - 10 febbraio [9751362], available at: https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9751362.
The Clearview case has also been referenced at the EU level in a news post by the European Data Protection Board (EDPB), available at: https://www.edpb.europa.eu/news/national-news/2022/facial-recognition-italian-sa-fines-clearview-ai-eur-20-million_en.

of this initiative is to ensure a correct functioning of the single market by fostering the development and deployment of trustworthy Artificial intelligence across the Union (European Union 2024).

The AI Act was designed as a comprehensive legislative framework that applies to all AI systems available or used within the EU. It is based on Article 114[12] and Article 16[13] of the Treaty on the Functioning of the European Union (TFEU). This framework aligns with the EUs strategy to ensure that products entering the EU market comply with existing legislation through conformity assessment and the CE marking system. However, due to its complexity, potential biases, and rapid evolution, AI often struggles to align with the legal requirements for certainty, transparency, accountability, and equal treatment. To address these challenges the EU legislator deliberately refrained from providing a rigid definition for "Artificial intelligence". Instead, the AI Act adopts a broad, nearly technology-neutral, and adaptable definition of "AI systems". In essence, the AI Act functions as a "Software Act". To reinforce this, the Commission incorporated a legal definition of "AI system" in Article 3 of the AI Act[14]. This definition encompasses various software-driven technologies that utilise specific methods and techniques, including "machine learning", "logic and knowledge-based" systems, and "statistical" methods. As stated in Article 2[15], the AI Act applies to providers who release or activate AI systems within the Union, regardless of whether they are based in the Union or outside of it; users of AI systems located within the Union and providers and users of AI systems located in a non-EU country when the AI system output is used within the Union (McCarthy 2007).

As a result, the AI Act does not extend to individual end-users, as the definition of a user in Art. 3(4)[16] explicitly excludes the use of AI systems in personal, non-professional contexts (Smuha *et al.* 2021). Accordingly, if the legislation can establish technical requirements for AI developers, it fails to establish robust and enforceable protection for fundamental rights. This happens due to weak provision, with several exceptions, aimed at safeguarding civic freedoms. Moreover, due to the need for flexibility in regulating evolving technologies, the final text of the Act contains several legal ambiguities and enforcement gaps, with many critical and specific aspects deferred to delegated acts, guidelines and future codes of conduct.

Nevertheless, the Commission has focused on adopting a risk-based regulation approach, grounded in the intended use of AI systems. Four categories of risk have been outlined[17]. At the top of this ranking system is the "unacceptable risk" category, which includes AI applications that undermine Union values and fundamental rights and are therefore prohibited. These prohibitions, listed in Article 5 (European Parliament and Council of the European Union 2024), include: subliminal techniques (resulting in physical and psychological harm); manipulative AI that leads to physical and psychological harm; social scoring and real-time remote biometric identification systems. The next category includes 'high-risk' AI technologies (Article 6), which are permitted but subject to compliance with specific AI requirements and ex-ante conformity assessments (Articles 16 and 43). Once these steps are successfully completed, the system can be affixed with the CE marking, signifying conformity with EU legislation. At this point, the AI application is ready to be introduced in the market or deployed for use. However, compliance does not end at market entry. A post-market monitoring procedure is also

12 Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:12008E114.

13 "Article 16 of the Treaty on the Functioning of the European Union grants all individuals the right to the protection of their personal data. It also requires the European Parliament and the Council of the European Union to lay down rules to protect individuals with regard to the processing of personal data by EU institutions, bodies, offices and agencies, and by the Member States when carrying out activities that fall within the scope of EU law". Retrieved from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=LEGISSUM:data_protection.

14 Definition: "Article 3 (1) of the AIA defines AI systems as software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with" (Ruschemeier 2023).

15 Available at: https://artificialintelligenceact.eu/article/2/.

16 Definition: "deployer means 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". Retrieved from: https://artificialintelligenceact.eu/article/3/.

17 AI Act as risk-based approach regulation.

required (Article 72[18]). Providers must establish and maintain a post-market monitoring system tailored to the characteristics of the AI technologies, and to the risks associated with the high-risk AI system throughout their entire lifecycle.

In this phase, two key requirements come into play: data governance and human oversight. The first provision (Article 13) addresses concerns in relation to the quality of data employed in the development of AI systems, thus including rules on how training datasets should be structured and utilised along with guidelines for data preparation. Human oversight requirements (Article 14) ensure that AI systems are designed and developed in a way that enables effective human supervision throughout their operation period, focusing on "empowering human overseer" to detect irregularities, "automation bias," interpreting the system's outputs, and identifying faults where necessary.

The ultimate category addressed in the AI Act is the "limited risk" and the "minimal or no risk" AI systems. The former must be subject to information and transparency obligations, as outlined in Article 52[19]. This provision obliges providers to notify users when they are interacting with an AI system, if not clearly evident. Additionally, it is mandatory to apply labels to deep fakes so that users can easily recognise them.

"Minimal or no risk" AI systems are permitted without restrictions. However, the European Commission encourages the voluntary adoption of codes of conduct for AI with specific transparency requirements.

The EU AI Act presents both opportunities and challenges for European AI developers. On the one hand, meeting all the aforementioned requirements and administrative burdens could hinder the market entry of a product, consequently stifling innovation and research in this sector. On the other hand, the AI requirements are necessary to guarantee trust and safety, allowing users to interact with AI systems that meet high standards of transparency and security. Furthermore, the AI Act aligns with the GDPR, increasing and balancing the need for a high level of safety. Where the GDPR imposes stringent regulations on personal data protection, the AI Act governs the development and use of AI systems, including cases where personal data is not involved. Ultimately, the Act also extends to non-European companies operating in the EU market.

As Europe continues its digital transformation, both frameworks will play a crucial role in promoting innovation.

U.S. legal approach to Artificial intelligence

As AI-related legislation progresses around the world, the United States legislative activity highlights the nation's dedication to adapting to Artificial intelligence. The U.S. does not have a unified AI Act. Instead, its approach consists of a collection of policies designed to boost innovation while managing associated risks. The most recent one was released by US President Joe Biden, whose Executive Order on the "Safe, Secure, and Trustworthy Development and Use of AI" was adopted on 30 October 2023[20]. This Order aims to address algorithmic discrimination and securing voluntary commitments from leading US companies (such as Amazon, Google, Meta, Microsoft, and OpenAI) to foster safe, secure, and trustworthy AI advancement. The Order addresses different policy areas. Firstly, it emphasises the need for new standards regarding AI safety and security. For example, it mandates that developers of AI systems share their safety testing results and other important data with the US government; it safeguards citizens'privacy against AI-related threats, emphasising federal efforts to hasten the adoption and utilisation of privacy-preserving methods; it promotes equity and civil rights to defend consumers, patients, and students. This is just the latest Executive Order on the matter, although legislative proposals have been emerging since 2020, when the National Artificial intelligence Initiative Act provided guidance for AI research, development, and evaluation activities within federal science agencies[21]. Additional acts have mandated specific agencies to lead AI programs and policies throughout the federal

18 Available at: https://artificialintelligenceact.eu/article/72/.

19 Available at: https://artificialintelligenceact.eu/article/52/.

20 As also recalled in a one-year-later press release by the U.S. Department of Commerce (2024), *"Commerce Marks One-Year Anniversary of Historic Biden-Harris Initiative", a*vailable at: https://www.commerce.gov/news/press-releases/2024/10/commerce-marks-one-year-anniversary-historic-biden-harris.

21 U.S. Congress (2020), *AI in Government Act of 2020* (H.R. 2575), available at: https://www.congress.gov/bill/116th-congress/house-bill/2575.

government, such as the "AI in Government Act"[22] and the "Advancing American AI Act"[23].

Nonetheless, with the commencement of Donald Trump's second term, many of the initiatives introduced by Biden are now being revoked. In fact, President Trump enacted a new Executive Order called "Removing Barriers to American Leadership in Artificial intelligence"[24] This directive aims to eliminate rules viewed as hindering AI innovation, facilitating an "unbiased and agenda-free" approach to the development of AI technologies.

At the state level, jurisdictions such as Colorado, Illinois, and California are actively shaping the legislative environment regarding AI compliance for businesses. As an example, the state of Colorado, inspired by the EU AI Act, adopted a risk-based legislation to regulate the use of high-risk AI systems, emphasising transparency, risk management, and thorough documentation throughout AI system development (Siegal and Garcia 2024).

3. A Comparison of EU and U.S. Approaches to AI Regulation

As the AI sector grows, so do the risks associated with AI technologies, making it more challenging to draft legislation that clearly defines measures, obligations, and boundaries for businesses in this field (Software Improvement Group 2024). European legislators have opted for a human-centric approach recognising the multiple and severe risks that Artificial intelligence systems pose to human rights, safety, the rule of law, and democracy (*ibidem*). In contrast, the U.S. approach prioritises the private sector, avoiding stringent AI regulation and thereby allowing businesses to expand in the sector.

The present section provides a comparative analysis of the European and American approaches to AI legislation, exploring their differences, similarities, strengths, and weaknesses.

The AI Act and the Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial intelligence present distinct approaches to the governance and regulation of AI. The AI Act is a regulation that applies directly and is legally binding in all EU Member States. In contrast, the Executive Order is a directive from the President of the United States that carries legal weight but influences policy and governance at the federal level without requesting Congressional approval.

When assessing the differences between the EU AI Act and the U.S. Executive Order, several immediate contrasts become apparent. The Executive Order primarily sets forth guidelines for federal agencies to follow, influencing industry practices. However, it does not impose regulations on private companies, except for specific reporting requirements under the Defense Production Act. On the other hand, the EU AI Act applies to any provider of general-purpose AI models operating within the EU, thereby having a much broader reach. While the U.S. Executive Order can be altered or revoked, especially with changes in presidential administrations, as seen with President Trump's election, the European AI Act establishes legally binding regulations within a more permanent governance framework. This difference highlights Europe's centrally coordinated and comprehensive regulatory framework, which can enact more binding rules compared to the U.S. approach. As noted, the U.S. Congress tends to leave more room for businesses to operate (Engler 2023). In contrast, the EU has placed a stronger emphasis on public transparency and on ensuring information about the role of AI in society. In this respect, the EU AI Act adopts a broader perspective on systemic risks, including issues such as widespread discrimination, significant accidents, and human rights implications. Both perspectives opted for a risk-based approach and similar principles in regulating AI, to ensure and develop a trustworthy AI system (*ibidem*).

Both legislative frameworks highlight the relevance of technical standards and best practices to achieve their goals. While the AI Act refers to them in Articles 17, 31, 32 and 40 – promoting a collaborative approach with European institutions and stakeholder organisations – the American Executive Order relies on federal agencies and expert bodies for implementation guidance, allowing the legal landscape to adapt over time. Therefore,

22 U.S. Congress (2020), *AI in Government Act of 2020* (H.R. 2575), available at: https://www.congress.gov/bill/116th-congress/house-bill/2575.

23 U.S. Congress (2021), *Advancing American AI Act* (S.1353), available at: https://www.congress.gov/bill/117th-congress/senate-bill/1353.

24 Available at: https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/.

this comparison is particularly relevant for global AI governance. By examining how two leading jurisdictions approach AI regulation, stakeholders can identify opportunities for dialogue, cooperation, and the development of shared norms. This process is vital not only to minimise regulatory conflicts but also to ensure that AI systems serve humanity in a safe, just, and inclusive manner.

Conclusion

Although Artificial intelligence seems to represent one of the greatest technological revolutions of our times due to its capability of transforming the way we interact with information, make decisions and manage critical processes, its implications extend far beyond scientific progress, raising fundamental questions related to human rights, transparency, and ethics. The issues analysed in this article – ranging from privacy concerns to discrimination and the spread of misinformation – demonstrate that the adoption of AI systems cannot occur without an adequate regulatory framework. The rapid development of AI has often outpaced lawmakers'ability to regulate its use, leading to scenarios where these technologies are deployed without sufficient safeguards for individuals. The case studies examined, including the risks associated with predictive algorithms in the judicial system and the proliferation of AI-generated content that distorts public discourse, highlight the urgent need for stronger regulatory intervention.

The adoption of the European Artificial intelligence Act has been a significant step in AI regulation, introducing a risk-based approach model capable of balancing technological innovation with the protection of fundamental rights, imposing stricter obligations on high-risk systems and banning unacceptable practices, such as social scoring and real-time biometric surveillance. The United States, on the other hand, has taken a more flexible and decentralised approach, with fragmented regulations across federal and state levels, influenced by the interests of the technology industry. The comparison between these two models highlights not only philosophical and legal differences but also the challenges of global AI governance.

Looking into the future, it seems to be necessary to develop an international regulatory ecosystem to coordinate the challenges posed by Artificial intelligence and ensure fairness, reliability, and security globally.

Moreover, scientific research must focus on ensuring AI systems' transparency and interpretability. In this respect, the notion of Trustworthy AI, promoted by the European Union, must translate into systems capable of explaining their decisions and operating responsibly. This implies the adoption of algorithmic auditing tools, increased accountability for AI providers, and stronger human oversight mechanisms (Laux *et al.* 2024).

This article has investigated the significant influence of Artificial intelligence on modern society. AI has progressed from its initial conception by McCarthy to the advanced neural networks used today. It represents a technological revolution due to its ability to analyse extensive datasets, identify intricate patterns, and automate decision-making processes that, in some tasks, surpass human performance. Nevertheless, this remarkable technological progress raises ethical and legal concerns regarding how AI systems function, how data is used, and the impact on individuals and communities. The analysis presented in the current article has highlighted the threats posed to essential human rights – particularly the right to privacy, access to reliable information, and protection against discrimination – providing clear examples of violations of these rights. The personalisation of algorithms driven by AI and deepfake technology contributes to the spread of misinformation, hindering individuals'access to accurate and diverse information. As previously stated, systems relying on automated decision-making can perpetuate biases, resulting in discriminatory practices in recruitment, law enforcement, and access to financial services. Additionally, AI surveillance tools, such as facial recognition, endanger privacy by facilitating extensive data gathering and intrusive monitoring. The risks highlighted emphasise the critical necessity for a regulatory framework that ensures AI development and implementation is with human rights standards.

The European Union is taking significant steps to mitigate the dangers arising from such technologies through initiatives like the AI Act, which specifically regulates high-risk AI applications while promoting transparency, accountability, and fairness within AI systems. Nonetheless, obstacles remain in finding a balance between innovation, ethical considerations

and legal obligations. Ultimately, in order to fully harness AÌs potential, it must be developed in a way that prioritises trust and protects fundamental rights. This raises the necessity of ensuring the robustness of AI systems, so that they continue to function as intended even when their inputs or models are subjected to adversarial interference. Assessing the robustness of a system requires testing for any possible input perturbations. However, this is practically impossible for AI, as the number of possible interferences is often exorbitantly large. If AI robustness cannot be adequately assessed, neither can its trustworthiness (Tocchetti *et al.* 2022).

Philosophical analyses define trust as the decision to delegate a task without any form of control or supervision over the way the task is executed. Successful instances of trust rest on an appropriate assessment of the trustworthiness of the agent to which the task is delegated (the trustee). Therefore, trustworthiness can be considered both a prediction of the likelihood that the trustee will behave as expected, based on past behaviour, and a measure of the risk borne by the trustor should the trustee fail to do so (Li *et al*. 2023). According to the International Organisation for Standardisation, trustworthy AI is "a framework to ensure that a system is worthy of being trusted based on the evidence concerning its stated requirements. It ensures that users'and stakeholders'expectations are met in a verifiable way (ISO/IEC 2020)".

The lack of transparency and the learning abilities of AI systems, along with the nature of attacks on these systems, make it difficult to evaluate whether the system will continue to behave as expected in any given context. Thus, it is important to de-anthropomorphise AI; AI systems cannot be held as responsible agents because no penalty imposed on them would motivate them to improve their behaviour (Li and Suh 2021). Policies should mandate a cautious evaluation of the level of control over AI systems based on the tasks they are designed to perform. In conclusion, the concept of reliability should be preferred over trustworthiness. A reliable AI system would require constant human oversight, ensuring that human operators understand how to approach different levels of automation. In particular, it is necessary to guard against the "half-automation problem": the phenomenon in which, when tasks are highly but not fully automated, the human operator tends to rely on the AI system as if it were fully automated. This underscores the importance of incorporating "kill switches" to maintain human control over AI processes, especially in the field of automated responses (Polito and Pupillo 2024).

The European Union has moved in this direction with the 2019 publication of the "Ethics Guidelines for Trustworthy Artificial intelligence," which established seven key requirements that AI systems should meet to be deemed trustworthy: human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination, fairness, societal and environmental well-being, and accountability. These guidelines urge all parties involved to adopt these seven essential criteria and to promote the Union's focus on human-centric principles (European Commission 2019). Nevertheless, to date, it remains vital to maintain reliable AI systems, thus ensuring human supervision to reduce risks and preserving authority over automated operations (Floridi 2021). An approach centred on human values, as encouraged by the EU, is key to aligning AI advancement with ethical and legal guidelines, cultivating trust and responsibility.

References

Amariles D.R., Baquero P.M. (2023), Promises and limits of law for a human-centric artificial intelligence. *Computer Law & Security Review*, *48*, 105795 <https://doi.org/10.1016/j.clsr.2023.105795>

Angwin J., Larson J., Mattu S., Kirchner L. (2016), *How we analyzed the COMPAS recidivism algorithm,* ProPublica <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>

Associated Press (2024), Company that sent fake Biden robocalls in New Hampshire agrees to $1m fine, *The Guardian*, 22 August

Blank I.A. (2023), What are large language models supposed to model?, *Trends in Cognitive Sciences*, 27, n.11, pp.987-989

Block H.D. (1962), The perceptron: A model for brain functioning, *Reviews of Modern Physics*, 34, n.1, 123-135

Bontridder N., Poullet Y. (2021), The role of artificial intelligence in disinformation, *Data&Policy*, 3, n.e32 <doi:10.1017/dap.2021.20>

Bozdag E. (2013), Bias in algorithmic filtering and personalization, *Ethics and information technology*, n.15, pp.209-227

Brackey A. (2019), *Analysis of racial bias in Northpointe's COMPAS algorithm*, Master's thesis at Tulane University https://library.search.tulane.edu/discovery/fulldisplay/alma9945513177106326/01TUL_INST:Tulane

Couldry N., Mejias U.A. (2019), Data colonialism: Rethinking big data's relation to the contemporary subject, *Television & New Media*, 20, n.4, pp.336-349

Council of Europe (2024), *Council of Europe Framework Convention on Artificial intelligence and Human Rights, Democracy and the Rule of Law*, Council of Europe Treaty Series n.225, Vilnius, Council of Europe

De Capitani Di Vimercati S., Foresti S., Livraga G., Samarati P. (2012), Data privacy: Definitions and techniques, *International Journal of Uncertainty*, *Fuzziness and Knowledge-Based Systems*, 20, n.06, pp.793-817

Deb E. (2023), COMPAS — an AI tool sending or keeping people in Jail, *edblogs.medium.com*, december 24

DeVries W.T. (2003), Protecting privacy in the digital age, *Berkeley Technology Law Journal*, 18, n.1, pp.283-311

Dul C. (2022), Facial recognition technology vs privacy: The case of Clearview AI, *The Queen Mary Law Journal*, 23, pp.1-24

Engler A. (2023), *The EU and U.S. diverge on AI regulation: A transatlantic comparison and steps to alignment, Brookings Institution*, United States of America, Retrieved from https://coilink.org/20.500.12592/3ptdkj on 21 Jul 2025

European Commission (2021), *Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial intelligence Act) and amending certain Union legislative acts* COM/2021/206 final

European Commission (2020), *White paper on artificial intelligence: A European approach to excellence and trust* COM(2020) 65 final

European Commission (2019), *Building trust in human-centric artificial intelligence,* COM(2019) 168 final

European Parliament, Council of the European Union (2024), *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 various regulations and directives (Artificial intelligence Act) (PE/24/2024/REV/1),* Official Journal of the European Union

European Parliament, Council of the European Union (2016), *Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). Official Journal of the European Union,* L 119, 1-88 http://data.europa.eu/eli/reg/2016/679/oj

European Union (2024), *Artificial intelligence Act*, Official Journal of the European Union, L 1689, 12 July https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

European Union Agency for Fundamental Rights (2022), *Bias in algorithms – Artificial intelligence and discrimination*, Luxembourg, Publications Office of the European Union

Festré A., Garrouste P. (2015), The 'Economics of Attention': A History of Economic Thought Perspective, *OEconomia,* 5, n.1, pp.3-36

Floridi L. (2021), The european legislation on ai: A brief analysis of its philosophical approach, *Philosophy & Technology*, 34, n.2, pp.215-222

Hardesty L. (2017), Explained: Neural networks. Ballyhooed artificial-intelligence technique known as "deep learning" revives 70-year-old idea, *MIT News,* April 14

Hsu T. (2024), New Hampshire Officials to Investigate A.I. Robocalls Mimicking Biden*, The New York Times*, January 22

ISO/IEC (2020*), TR24028:2020 Information Technology – Artificial intelligence – Overview of Trustworthiness in Artificial intelligence [Standard],* Vernier (Geneva), International Organization for Standardization

Kenaw S. (2008), Hubert l. Dreyfus's critique of classical ai and its rationalist assumptions, *Minds and Machines*, 18, n.2, pp.227-238

Kertysova K. (2018), Artificial intelligence and Disinformation: How AI Changes the Way Disinformation is Produced, Disseminated, and Can Be Countered, *Security and Human Rights,* 29, n.1-4, p.55-81

Laux J., Wachter S., Mittelstadt B. (2024), Trustworthy artificial intelligence and the European Union AI Act: On the conflation of trustworthiness and acceptability of risk, *Regulation & Governance*, 18, n.1, pp.3-32

LeCun Y., Bengio Y., Hinton G. (2015), Deep learning, *Nature*, 521, n.7553, pp.436-444 https://doi.org/10.1038/nature14539

Li B., Qi P., Liu B., Di S., Liu J., Pei J., Yi J., Zhou B. (2023), Trustworthy ai: From principles to practices, *ACM Computing Surveys*, 55, n.9, pp.1-46

Li M., Suh A. (2021), Machinelike or humanlike? A literature review of anthropomorphism in AI-enabled technology, in HICCS, *Proceedings of the 54th Hawaii international conference on system sciences*, Kauai (Hawaii, USA), ScholarSpace, pp.4053-4062

Madiega T. (2021), *Artificial intelligence act,* Bruxelles, European Parliamentary Research Service

Mantelero A. (2022), *Beyond data: Human rights, ethical and social impact assessment in AI*, London, Springer Nature

McCarthy J. (2022), Artificial intelligence, logic, and formalising common sense, in Carta S. (ed.), *Machine Learning and the City: Applications in Architecture and Urban Design*, Bognor Regis (UK), John Wiley & Sons Ltd., pp.69-90

McCarthy J. (2007), *What is Artificial intelligence*, Stanford (CA), Stanford University

Neapolitan R.E. (2012), *Contemporary Artificial intelligence*, Boca Raton (USA), Chapman and Hall/CRC

Nelson G.S. (2019), *Bias in artificial intelligence*, *North Carolina medical journal*, 80, n.4, pp.220-222

Pamment J. (2020), *The EU Code of Practice on Disinformation: Briefing Note for the New EU Commission*, Policy perspective series n.1, Washington (DC), Carnegie Endowment for International Peace

Pisanu N. (2022), Il caso Clearview AI, il riconoscimento facciale viola la nostra privacy: la multa del Garante, *Cybersecurity 360*, 9 marzo

Polito C., Pupillo L. (2024), Artificial intelligence and cybersecurity, *Intereconomics,* 59, n.1, pp.10-13

Rombach R., Blattmann A., Lorenz D., Esser P., Ommer B. (2022), High-resolution image synthesis with latent diffusion models, in *CVPR 2022. 2022 Conference on Computer Vision and Pattern Recognition (CVPR)*, Danvers (MA), Clearance Center, pp.10674-10685

Rosenblatt F. (1958), The perceptron: A probabilistic model for information storage and organization in the brain, *Psychological Review*, 65, n.6, pp.386-408

Ruschemeier H. (2023), AI as a challenge for legal regulation–the scope of application of the artificial intelligence act proposal, *Era Forum*, 23, n.3, pp. 361-376

Seow J.W., Lim M.K., Phan R. C.W., Liu J.K. (2022), A comprehensive overview of Deepfake: Generation, detection, datasets, and opportunities, *Neurocomputing*, 53, pp.351-371

Siegal A., Garcia I. (2024), A deep dive into Colorado's artificial intelligence act, *Attorney general Journal*, October 26

Sohail S., Farhat F., Himeur Y., Nadeem M., Madsen D.Ø., Singh Y., Atalla S., Mansoor W. (2023), *The Future of GPT: A Taxonomy of Existing ChatGPT Research, Current Challenges, and Possible Future Directions*, 8 April <http://dx.doi.org/10.2139/ssrn.4413921>

Smolensky P. (1987), Connectionist AI, symbolic AI, and the brain, *Artificial intelligence Review*, 1, n.2, pp.95-109

Smuha N., Ahmed-Rengersb E., Harkens A., Li W., MacLaren J., Piselli R., Yeung K. (2021), *How the EU can achieve legally trustworthy AI: a response to the European Commission's Proposal for an Artificial intelligence Act*, Birmingam, LEADS Lab @ University of Birmingam <http://dx.doi.org/10.2139/ssrn.3899991>

Software Improvement Group (2024), *AI legislation in the US: A 2025 overview* https://www.softwareimprovementgroup.com/us-ai-legislation-overview/

Tocchetti A., Corti L., Balayn A., Yurrita M., Lippmann P., Brambilla M., Yang J. (2022), Ai robustness: a human-centered perspective on technological challenges and opportunities, *ACM Computing Surveys*, 57, n.6, Article 141

Treusch P. (2020), Re-reading ELIZA: Human–machine Interaction as cognitive sense-ability, in Lorenz-Meyer D., Treusch P., Liu X. (eds.), Feminist Technoecologies. *Reimagining Matters of Care and Sustainability,* London, Routledge, pp.61-76

Turillazzi A., Taddeo M., Floridi L., Casolari F. (2023), The digital services act: an analysis of its ethical, legal, and social implications, *Law, Innovation and Technology*, 15, n.1, pp. 83-106

Turing A.M. (1950), Computing Machinery and Intelligence, *Mind*, n.49, pp.433-460

Waldemarsson C. (2020), *Disinformation, Deepfakes & Democracy: The European response to election interference in the digital age*, Copenhagen, The Alliance of Democracies Foundation

West D.M. (2018), What is artificial intelligence?, *Brookings.edu*, October 4

Xu D., Fan S., Kankanhalli M. (2023), Combating misinformation in the era of generative ai models, in *Proceedings of the 31st ACM International Conference on Multimedia*, New York, Association for Computing Machinery, pp.9291-9298

Laura Evangelista
l.evangelista@inapp.gov.it
Researcher, Head of the Research Group Accreditation and Quality for VET at INAPP. National Coordinator of the EQAVET National Reference Point for Italy. She coordinates research projects related to the monitoring and evaluation of ccreditation systems of VET providers in Italy, the implementation of quality arrangements in line with EU-evel principles and indicators, the experimentation of Peer Review at provider and system level, and the dissemination of results through seminars, conferences and papers. Recent publications: (with Fonzo C.), Self-assessment in VET and Higher education: links and further developments, *Quaderni di Comunità*, n.3, 2023; (with Fonzo C.), La metodologia europea della Peer Review: prima sperimentazione tra istituti scolastici e Centri di Formazione Professionale, *Rassegna CNOS*, n.39, 2023.

Concetta Fonzo
c.fonzo@inapp.gov.it
Deputy Coordinator of the EQAVET National Reference Point for Italy and Deputy Coordinator of ReferNet Italy. Member of INAPP's research group Accreditation and Quality for Training and of its National Operational Programmes funded by the ESF+. Since 2004, she has been a member of scientific and technical committees and boards for various European initiatives and networks. An expert in European project management, she contributes to national and international research activities in the fields of education, training, guidance and social inclusion. Recent publications: (with Evangelista L.), Self-assessment in VET and Higher education: links and further developments, Quaderni di Comunità, n.3, 2023; (with Evangelista L.), La metodologia europea della Peer Review: prima sperimentazione tra istituti scolastici e Centri di Formazione Professionale, *Rassegna CNOS*, n.39, 2023.

Eleonora Zecca
ele.zecca00@gmail.com
She holds a Master's degree in Policies and Governance in Europe from LUISS Guido Carli and a Bachelor's degree in Political Science and International Relations from the University of Parma. She also studied at the University of Passau (Germany) as part of a double degree experience. Her research interests focus on labor market dynamics and political systems. She authored her Master's thesis on *The Italian Labor Market and the Green Transition in SMEs. A Focus on the Automotive Sector* and her Bachelor's thesis on *Polarizzazione affettiva e sistemi di partito. Un'analisi comparata*.