World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(REVIEW ARTICLE)

Check for updates

Safeguarding human dignity: A narrative review of prohibited AI practices under the EU AI Act

Dinesh Deckker [1,*] and Subhashini Sumanasekara [2]

[1] *Department of Science and Technology, Wrexham University, United Kingdom.*
[2] *Department of Computing and Social Sciences, University of Gloucestershire, United Kingdom.*

World Journal of Advanced Research and Reviews, 2025, 26(03), 243–260

Publication history: Received on 28 April 2025; revised on 31 May 2025; accepted on 03 June 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.26.3.2193

Abstract

Artificial intelligence (AI) technologies are reshaping public administration, law enforcement, and social governance—but not without raising profound human rights concerns. This narrative review examines the eight AI practices explicitly prohibited under Article 5 of the European Union Artificial Intelligence Act (EU AI Act), which categorizes them as presenting an "unacceptable risk." These practices include subliminal manipulation, exploitation of vulnerable populations, social scoring, predictive profiling, untargeted scraping of biometric data, emotion recognition in sensitive settings, biometric categorization by sensitive attributes, and real-time biometric surveillance in public spaces. The purpose of this review is to assess how each prohibition corresponds to specific human rights protections, such as autonomy, privacy, non-discrimination, and dignity, and to explore the legal and ethical frameworks that justify such prohibitions.

This study employs a qualitative narrative methodology, integrating legal analysis, historical misuse cases, and ethical theory—drawing from sources including the EU Charter of Fundamental Rights, the European Convention on Human Rights, and scholarly work in AI ethics. Key findings reveal that each prohibited AI practice has precedent in past abuses and can be normatively justified using deontological, utilitarian, and virtue ethics frameworks.

The review concludes that the prohibited AI systems not only breach legal standards but undermine the moral foundations of democratic societies. These findings support the necessity of rights-based AI regulation and underscore the EU's global leadership in normative governance. Future research should focus on enforcement challenges, international harmonization, and the development of new safeguards for emerging AI risks.

Keywords: EU AI Act; Article 5; Human rights; Prohibited AI practices; Ethics; Biometric surveillance; Social scoring

1. Introduction

1.1. Purpose of the Review

Artificial intelligence (AI) systems are playing an increasingly influential role in shaping personal lives, social relations, and democratic governance. While the benefits of AI in terms of innovation, efficiency, and societal advancement are widely recognised, there is growing concern about the risks these systems pose, particularly when they infringe upon fundamental human rights. In response, Article 5 of the European Union Artificial Intelligence Act (EU AI Act) identifies a set of AI practices deemed to carry an unacceptable level of risk and, therefore, subject to prohibition.

* Corresponding author: Dinesh Deckker; ORCID - 0009-0003-9968-5934

This review examines the full scope of prohibited practices under Article 5, which include: subliminal manipulation; exploitation of vulnerable individuals based on age, disability, or socio-economic status; social scoring systems that assess individuals based on behaviour or personality traits; predictive risk assessments for criminal behaviour based solely on profiling; untargeted scraping of facial images to build biometric databases; emotion recognition in workplaces and educational institutions; biometric categorisation based on sensitive attributes such as race or religion; and the use of real-time remote biometric identification systems in public spaces for law enforcement purposes, except under narrowly defined conditions. The primary aim of this review is to analyse how each of these practices violates core human rights—namely, the rights to privacy, autonomy, non-discrimination, and human dignity—while situating these risks within broader legal, ethical, and societal frameworks.

Artificial intelligence (AI) introduces transformative possibilities, but it also brings with it complex regulatory and ethical challenges. Policymakers, industry leaders, and civil society actors across the globe are actively engaged in determining how to regulate AI in ways that safeguard fundamental rights without impeding technological innovation. The European Union's Artificial Intelligence Act marks a pivotal step in this direction, offering the first comprehensive risk-based legal framework for AI governance (Veale and Borgesius, 2021). Central to this framework is Article 5, which outlines AI systems considered to pose an "unacceptable risk"—technologies that are fundamentally incompatible with EU values and human rights protections.

Global apprehension regarding the misuse of AI is growing, with academic and policy literature documenting numerous concerns. Key issues include manipulative techniques in digital advertising, discriminatory outcomes in algorithmic decision-making, and the expansion of biometric surveillance infrastructures (Zuboff, 2019; Eubanks, 2018). These developments underscore the urgent need to critically examine how the most harmful AI practices are being addressed within legal and ethical norms.

This review focuses on the prohibited AI practices outlined in Article 5 of the EU AI Act (2024), with a particular emphasis on their implications for ethics, law, and human rights. Specifically, it examines eight key categories:

- AI systems that employ subliminal techniques to influence individuals beyond their conscious awareness in harmful ways.
- AI systems that exploit individuals based on vulnerability related to age, disability, or socio-economic status.
- AI systems used for social scoring by public entities, leading to unjust or disproportionate treatment.
- AI systems that make risk assessments to predict criminal behavior based solely on profiling or inferred personality traits.
- AI systems that create or expand facial recognition databases through untargeted scraping of images from the internet or CCTV.
- AI systems that infer emotions in workplaces or educational institutions, unless justified for safety or medical purposes.
- Biometric categorization systems that deduce sensitive attributes such as race, religion, or sexual orientation.
- Real-time remote biometric identification systems deployed in public spaces by law enforcement, except under strict legal exceptions.

The review aims to assess the human rights risks associated with these practices and situate them within a broader historical, legal, and normative context.

1.2. Importance of the Topic

The prohibition of specific AI applications is not merely a technical matter—it is a crucial ethical and human rights obligation. When manipulative or discriminatory AI systems operate without regulation, they can undermine democratic values, erode protections for vulnerable populations, and normalise practices such as surveillance, coercion, and systemic inequality. Understanding the rationale behind these prohibitions is crucial for shaping future legislation, informing international policy efforts, and fostering the ethical development of AI technologies.

Ethicists and scholars have long warned of the risks associated with AI, especially regarding privacy (Tufekci, 2015), surveillance capitalism (Zuboff, 2019), and data-driven discrimination (Eubanks, 2018). Floridi et al. (2018) have articulated governance principles grounded in human dignity, autonomy, and justice. Global instruments such as the UNESCO Recommendation on the Ethics of AI (2021) and the OECD AI Principles (2019) further underscore the importance of rights-based approaches. However, a dedicated analysis of the explicitly prohibited AI practices within the EU legal framework—and their specific human rights foundations—remains an area requiring deeper scholarly attention.

1.3. Research Gap

Although there is extensive discourse on ethical AI and frameworks for responsible innovation, limited scholarship has directly linked the provisions of Article 5 of the EU AI Act to concrete human rights violations. Existing literature tends to address ethical concerns in general terms, without systematically analysing how each prohibited AI practice correlates with specific rights infringements. Furthermore, there is a noticeable lack of integrated analysis that situates these prohibitions within the context of historical misuse cases, relevant legal precedents, and the ethical theories that substantiate their necessity.

1.4. Research Aim

This study aims to:

- Examine how each prohibited AI practice in Article 5 contravenes fundamental rights.
- Situate the prohibitions within relevant legal, historical, and ethical contexts.
- Assess potential gaps between regulatory intent and practical enforcement.

1.5. Research Questions

- How does each prohibited AI practice under Article 5 threaten specific human rights?
- What legal and ethical frameworks justify the prohibition of these AI systems?
- What real-world examples illustrate the risks of allowing these systems?
- Are there AI systems currently in use that approach or cross these ethical boundaries?

1.6. Main Contributions

This review offers:

- A human rights-based framework for understanding the EU's "unacceptable risk" classification.
- A synthesis of ethical and legal justifications for AI prohibitions.
- An evaluation of contemporary AI systems that may risk violating Article 5.

2. Theoretical and Legal Background

2.1. Human Rights Foundations

The prohibition of specific artificial intelligence (AI) practices under Article 5 of the EU Artificial Intelligence Act (AI Act)

2.1.1. UN Guiding Principles on Business and Human Rights

Adopted in 2011, the UN Guiding Principles on Business and Human Rights (UNGPs) articulate the duty of both states and corporations to protect, respect, and remedy human rights violations in the context of commercial activity. These principles underline the obligation of private sector participants—specifically, AI developers and implementers—to protect rights including non-discrimination, privacy, and autonomy (United Nations Human Rights Council, 2011). Under the UNGPs, companies are required to conduct human rights due diligence, a principle also reflected in the EU AI Act's requirement for risk management systems and impact assessments in high-risk contexts. Although Article 5 focuses on the most severe cases, its essence resonates with the overarching UNGP philosophy of actively preventing and mitigating harm.

2.1.2. UNESCO's 2021 Recommendation on the Ethics of AI

In a more recent development, the UNESCO Recommendation on the Ethics of Artificial Intelligence (2021) outlines a global normative framework for the governance of AI. The recommendation emphasises four core values: respect for human rights, human dignity, environmental sustainability, and peace. The principles of proportionality and not harm are fundamental, requiring that AI systems not only steer clear of causing harm but also protect human agency and promote inclusion (UNESCO, 2021). These principles support a ban on manipulative or discriminatory AI technologies, as they fail to meet ethical thresholds of fairness, accountability, and transparency.

2.2. Ethical Theories in AI

The regulatory prohibitions outlined in Article 5 of the EU AI Act reflect deep-seated ethical concerns about the potential misuse of artificial intelligence. A normative analysis of these prohibited practices benefits from a grounding in ethical theories that inform both moral reasoning and policy-making. This section outlines how deontology, utilitarianism, and virtue ethics provide philosophical justification for restricting specific AI systems. It also highlights contemporary contributions from leading scholars in AI ethics, particularly Luciano Floridi, Brent Mittelstadt, Reuben Binns, and James Moor, whose work has shaped the digital ethical landscape.

2.2.1. Deontology: Autonomy and Moral Duties

Deontological ethics, particularly Immanuel Kant's philosophy, prioritises moral duties and respect for autonomy. From this standpoint, individuals must always be treated as ends in themselves, not merely as means to an end. AI systems that employ subliminal techniques to manipulate behaviour, as banned under Article 5(1)(a), fundamentally violate this principle, as they bypass rational agency and undermine individual consent (Kant, 1785/1996). Kantian ethics also holds that actions are morally impermissible if they cannot be universally accepted. Thus, exploitative AI that targets vulnerable persons (Article 5(1)(b)) fails to uphold a duty of moral protection owed to those with diminished capacity for resistance or choice.

Floridi (2013) builds upon deontological principles in his concept of "inforgs"—informational organisms that exist within a shared infosphere—and argues for preserving informational dignity, particularly in contexts of surveillance and data manipulation. Respecting informational autonomy in AI design becomes not only a moral requirement but a structural condition for ethical digital environments.

2.2.2. Utilitarianism: Assessing Harms and Benefits

Utilitarian ethics evaluates actions based on their consequences—maximising benefits and minimising harms for the greatest good. While seemingly tolerant of cost-benefit trade-offs, utilitarianism also offers a robust critique of high-risk or discriminatory AI systems, primarily when the harms are unequally distributed. For example, social scoring mechanisms (Article 5(1)(c)) can yield reputational harm, restricted access to public services, and unjust discrimination—outcomes that disproportionately affect marginalized individuals without yielding proportionate societal benefits (Taddeo and Floridi, 2018). The utilitarian calculus, when fairly applied, would deem such systems ethically indefensible.

Moreover, biometric surveillance (Article 5(1)(d)) in public spaces may promise public safety but introduces widespread chilling effects, normalises suspicion, and erodes trust—collective harms that arguably outweigh any purported security gains (Wright and Raab, 2012). Mittelstadt (2017) notes that while AI systems may be justified in their design, their social embedding often produces unintended negative externalities that utilitarian ethics must take into account.

2.2.3. Virtue Ethics: Character, Intent, and Social Well-being

Unlike rule-based or consequence-based approaches, virtue ethics emphasises moral character, intent, and the cultivation of human flourishing. Originating from Aristotle's philosophy, this framework judges actions not just by their rules or outcomes but by the virtues—or vices—they express. AI systems that manipulate or discriminate are thus problematic not only because of their impacts or legality, but also because they reflect a technological culture devoid of virtues such as honesty, compassion, and justice (Binns, 2014).

Virtue ethics also aligns with the precautionary moral stance often adopted in AI governance: even if an AI system *can* be developed, the question is whether a prudent, wise, and just society *should* deploy it. Moor (2006) echoes this by proposing that ethical governance of AI must incorporate virtues such as responsibility, transparency, and accountability—virtues sorely lacking in the cases targeted by Article 5.

Table 1. Ethical Theories Justification Matrix for Prohibited AI Practices under Article 5 of the EU AI Act

Prohibited AI Practice	Deontology	Utilitarianism	Virtue Ethics
Subliminal Manipulation	Violates autonomy and consent	Creates disproportionate harm vs. benefit	Manipulative and dishonest design
Exploitation of Vulnerable Groups	Fails in their duty to protect the vulnerable	Unequal harm to high-risk groups	Lacks compassion and fairness
Social Scoring	Undermines dignity and equality	Social penalties outweigh societal gains	Promotes social shame and exclusion
Predictive Risk Profiling	Breaches the presumption of innocence	Leads to unjust targeting without evidence	Expresses distrust and prejudice
Untargeted Biometric Scraping	Lacks informed consent	Harms outweigh predictive utility	Disrespects personal identity
Emotion Recognition	Covert intrusion into mental states	No proven benefit; risk of systemic bias	Lacks humility and moral restraint
Biometric Categorization	Essentialises identity	Leads to unjust consequences	Encourages biased social judgment
Real-Time Biometric Surveillance	Violates privacy and individual rights	Widespread fear limits public good	Inhibits civic participation and courage

While distinct, these ethical theories converge on a shared judgment: the practices prohibited under Article 5 of the AI Act are ethically impermissible. Deontology decries the violation of autonomy; utilitarianism reveals disproportionate harms; virtue ethics condemns their morally corruptive nature. Contemporary theorists, such as Floridi, Moor, Mittelstadt, and Binns, have expanded these frameworks into actionable digital ethics, providing normative clarity that supports the legislative intent of the AI Act.

2.3. The EU Artificial Intelligence Act: Overview

2.3.1. Structure and Purpose of the AI Act

The European Union Artificial Intelligence Act (EU AI Act), proposed by the European Commission and formally adopted in 2024, represents the world's first comprehensive legal framework for regulating artificial intelligence (European Commission, 2024). The Act adopts a risk-based approach, categorising AI systems into four distinct levels: minimal risk, limited risk, high risk, and unacceptable risk. This classification system aims to strike a balance between innovation and the protection of fundamental rights and democratic values (Veale and Borgesius, 2021).

The overall aim of the AI Act is to establish legal certainty and foster trust in AI by ensuring that high-risk applications undergo rigorous conformity assessments, meet transparency obligations, and are subject to human oversight. The Act also aims to prevent the misuse of AI in contexts where human dignity, autonomy, and equality could be compromised (Floridi et al., 2018). Rather than stifling technological development, the Act frames regulation as a necessary scaffold to steer AI innovation toward ethical and socially beneficial outcomes.

2.3.2. Article 5: Defining "Unacceptable Risk"

At the core of the EU AI Act lies Article 5, which delineates eight specific AI practices classified as presenting an "unacceptable risk" and are therefore banned within the European Union. These include:

- AI systems using subliminal or manipulative techniques that significantly distort human behavior (Article 5(1)(a));
- Exploitation of vulnerable groups based on age, disability, or socio-economic status (Article 5(1)(b));
- Social scoring systems that result in unjustified or disproportionate treatment (Article 5(1)(c));
- Predictive policing AI systems based solely on personality profiling without objective evidence (Article 5(1)(d));
- Untargeted scraping of biometric data to build facial recognition databases (Article 5(1)(e));
- Emotion recognition systems in workplaces and schools (Article 5(1)(f));

- Biometric categorization systems that infer sensitive attributes such as race or sexual orientation (Article 5(1)(g));
- Real-time remote biometric identification in public spaces by law enforcement, except under tightly regulated exceptions (Article 5(1)(h)).

These prohibitions are not arbitrary; they are rooted in the European Charter of Fundamental Rights (CFR), specifically Articles 1 (respect for human dignity), 7 (respect for private and family life), and 8 (protection of personal data). As Taddeo and Floridi (2018) argue, framing specific AI applications as intrinsically harmful reflects a normative commitment to safeguard autonomy and prevent systemic discrimination.

2.3.3. Global Relevance and Legal Precedent

While the AI Act is EU-specific, its implications are global. It has the potential to set a de facto international standard, much like the General Data Protection Regulation (GDPR) influenced global privacy laws. Non-EU companies operating within the EU will be required to comply with the AI Act, creating a ripple effect in international legal and corporate practices (Cihon et al., 2021).

The prohibitions in Article 5 are also supported by jurisprudence from European and international courts. For instance, the European Court of Human Rights (ECtHR) has ruled in cases such as S. and Marper v. United Kingdom that the retention of biometric data can violate privacy rights under Article 8 of the European Convention on Human Rights (ECtHR, 2008). Similarly, the Digital Rights Ireland case invalidated mass data retention laws for being disproportionate, laying the groundwork for restrictions on surveillance-oriented AI (CJEU, 2014).

These precedents reinforce the EU's position that specific AI applications, particularly those involving opaque data processing, biometric surveillance, and discriminatory profiling, are not only unethical but also legally indefensible. As global debates on AI ethics intensify, the EU AI Act—particularly Article 5—serves as both a policy benchmark and a normative statement about the kinds of AI the world should reject.

3. Analysis of Prohibited Practices concerning Human Rights

3.1. Subliminal Manipulation

3.1.1. Defining Subliminal Manipulation in AI Contexts

Subliminal manipulation refers to the use of stimuli or processes below the threshold of conscious awareness to influence individuals' behaviour, preferences, or decisions without their knowledge. In the context of artificial intelligence, this manipulation is operationalised through algorithmic profiling, behavioural targeting, and affective computing designed to bypass rational deliberation and exploit unconscious biases (Zuboff, 2019). Article 5(1)(a) of the EU Artificial Intelligence Act explicitly prohibits AI systems that deploy such techniques "beyond a person's consciousness" and in a manner that causes or is likely to cause physical or psychological harm (European Commission, 2021).

While subliminal techniques are not new, AI systems significantly amplify their reach and precision through continuous data collection, psychographic profiling, and adaptive feedback loops. The risk escalates when these systems interact with vulnerable populations, such as children or individuals with mental health conditions, who may be less capable of recognising or resisting such influences (Susser et al., 2019).

3.1.2. Examples: AI-Powered Advertising and Affective Nudging

A well-documented domain of subliminal AI manipulation is behavioural advertising, where machine learning algorithms track users' digital footprints—clicks, pauses, scrolls—and personalise content that subtly nudges consumer behaviour. The infamous Cambridge Analytica scandal demonstrated how psychometric profiling of Facebook users was used to influence voting behaviour through emotionally charged microtargeted ads (Isaak and Hanna, 2018). Though not strictly "subliminal" in the classical sense, these techniques blur the line between persuasion and manipulation, particularly when users are unaware of the mechanisms guiding their choices.

Another emerging area is affective computing, where AI models analyse facial expressions, tone of voice, or physiological data to infer emotional states and adapt interactions accordingly. Such systems are increasingly used in customer service, recruitment, and even education, raising concerns about emotional exploitation and consent

(Crawford, 2021). The lack of transparency and explainability in these models further undermines users' ability to recognise manipulation, violating principles of informed consent.

3.1.3. Human Rights at Risk

The use of subliminal manipulation by AI systems directly threatens mental autonomy, a foundational concept in human rights and moral philosophy. According to Kantian ethics, autonomy is the capacity to act according to rational will, free from coercion or deception (Kant, 1785/1996). By circumventing conscious reasoning, subliminal AI systems treat individuals as mere means to behavioural ends—whether commercial, political, or social—which constitutes a moral and legal violation of human dignity (Floridi, 2013).

The right to dignity, enshrined in Article 1 of the Charter of Fundamental Rights of the European Union, is closely tied to the right to freedom of thought and psychological integrity (European Union, 2012). Subliminal AI undermines these protections by introducing covert influences that escape critical scrutiny. As Susser et al. (2019) argue, such systems infringe upon individuals' "mental self-determination," a right increasingly recognised as vital in the digital age.

Moreover, the absence of meaningful consent mechanisms in many AI-powered platforms violates data protection principles under the General Data Protection Regulation (GDPR), especially those requiring transparency and freely given, informed consent (Mantelero, 2018). In cases where manipulation targets marginalized or cognitively impaired individuals, the discriminatory impact compounds the ethical breach.

3.2. Exploitation of Vulnerable Groups

3.2.1. Targeting the Vulnerable: A Critical Human Rights Concern

Under Article 5(1)(b) of the EU Artificial Intelligence Act (AI Act), AI systems that exploit the vulnerabilities of individuals due to age, disability, or social or economic circumstances are classified as posing an unacceptable risk and are thus explicitly prohibited. The inclusion of this clause reflects a broader commitment to upholding the principles of equity, fairness, and non-discrimination within the digital ecosystem. This provision directly links to the fundamental rights protections enshrined in the EU Charter of Fundamental Rights, including the right to dignity (Article 1), non-discrimination (Article 21), and the rights of children and the elderly (Articles 24 and 25) (European Union, 2012).

AI systems deployed in contexts such as education, health, and welfare disproportionately affect children, persons with disabilities, the elderly, and economically disadvantaged communities—populations often lacking the resources, digital literacy, or institutional support to recognize or resist algorithmic harms (Eubanks, 2018). As such, the use of AI in these domains raises pressing ethical and legal concerns about power asymmetry, consent, and informed agency.

3.2.2. Examples of Exploitation

One domain where such exploitation manifests is in AI-powered educational tools that profile children based on test scores, behavioural data, or biometric responses to adapt instruction or determine learning outcomes. Without transparent oversight and accountability, these systems may reinforce biases, impose developmental ceilings, or cause psychological harm through labelling and surveillance (Crawford, 2021).

Similarly, in welfare systems, AI has been used to predict potential fraud or misuse of public benefits. In the Netherlands, the SyRI system (System Risk Indication) used algorithmic profiling to flag "at-risk" individuals for fraud investigations, disproportionately targeting low-income neighbourhoods with limited due process or transparency (Lepri et al., 2018). The Dutch courts ultimately ruled this practice discriminatory and unlawful, citing violations of privacy and human dignity (Allen and Masters, 2020).

Healthcare is another sensitive area, particularly with AI-driven triage or diagnostic tools that may disadvantage elderly or disabled individuals based on generalisations or biased training data. These populations often exhibit complex, non-normative patterns that resist statistical simplification, thereby increasing the risk of both exclusion and misclassification (Whittlestone et al., 2019).

3.2.3. Human Rights and Ethical Failures

At its core, the exploitation of vulnerable groups by AI represents an ethical failure to uphold the principles of equity and fairness, two pillars of both democratic governance and international human rights law. Fairness in AI is not simply a matter of equal performance across groups; it requires procedural justice, contextual sensitivity, and proactive inclusion in system design (Binns, 2018).

From a rights-based perspective, these AI systems violate the principle of non-discrimination, primarily when algorithmic decisions produce disparate impacts on protected groups. Moreover, the absence of meaningful human oversight in high-stakes decision-making settings undermines autonomy and accountability, contravening international guidelines such as the UN Convention on the Rights of Persons with Disabilities and the UN Convention on the Rights of the Child.

Floridi and Cowls (2019) argue that ethical AI must be designed to protect not only rights but also the capabilities necessary for human flourishing, especially for those who are structurally disadvantaged. Systems that exploit such groups for surveillance, control, or behavioural prediction erode human dignity and reinforce systemic injustice under the guise of technological neutrality.

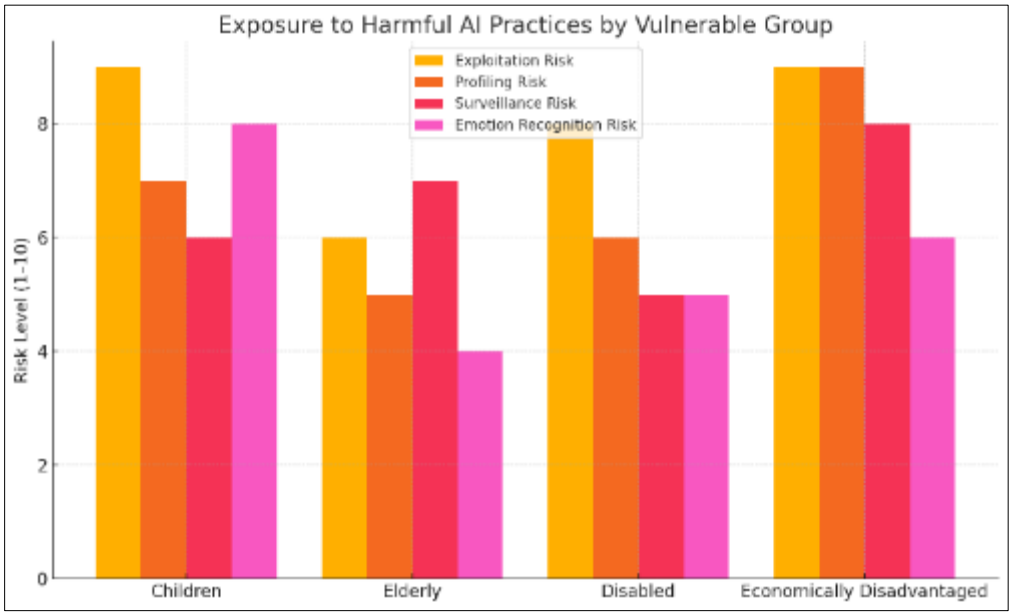

Figure 1 Digital Divide and Vulnerability Index: Comparative Exposure to Harmful AI Systems Across Vulnerable Groups

This interpretive index illustrates the relative exposure of different vulnerable populations—children, elderly individuals, persons with disabilities, and low-income communities—to high-risk AI practices prohibited under Article 5 of the EU AI Act. Scores reflect comparative assessments based on thematic analysis of existing academic literature, including Eubanks (2018), Zuboff (2019), Wachter et al. (2021), and Floridi et al. (2018). The index does not represent empirical measurement but serves as a heuristic tool to highlight structural risk disparities in AI deployment contexts such as welfare, education, surveillance, and profiling. Readers are advised to interpret the data as indicative rather than statistically validated.

3.3. Social Scoring

3.3.1. The Architecture of Algorithmic Judgment

Social scoring systems are AI-driven mechanisms that assign behavioural or reputational scores to individuals or groups based on observed or inferred activities. These scores are then used to determine access to public services, employment opportunities, travel freedoms, or even social recognition. Under Article 5(1)(c) of the EU Artificial Intelligence Act, such systems are explicitly prohibited when implemented by public authorities and when they lead to detrimental or unfavourable treatment in contexts unrelated to the behaviour assessed or that are unjustified or disproportionate (European Commission, 2021).

The prohibition on social scoring systems stems from the profound ethical and legal concerns associated with evaluating individuals based on their behaviour, characteristics, or perceived trustworthiness. Such systems risk institutionalizing bias, infringing on individual autonomy, and perpetuating discriminatory treatment. When AI systems are used to classify people based on opaque metrics, the result may be unequal access to services, stigmatization, and violations of due process. The EU explicitly outlaws such practices when they lead to unjustified or disproportionate treatment across contexts unrelated to the data's origin, reinforcing its commitment to human dignity and non-discrimination

(European Commission, 2024; Veale and Borgesius, 2021). By codifying this ban, the EU sets a legal precedent that prioritizes fairness, transparency, and fundamental rights in the deployment of AI.

3.3.2. Violations of Fundamental Rights

Social scoring systems inherently violate several fundamental rights and legal principles enshrined in both EU and international human rights law. First, they undermine the principles of equality and non-discrimination, as individuals may receive unequal treatment based on aggregated behavioural data that is often context-insensitive, opaque, and prone to error. The use of such data for decisions related to employment, mobility, or access to services risks reinforcing structural biases, particularly against socioeconomically disadvantaged or politically marginalised groups (Wachter et al., 2021).

Second, these systems challenge the presumption of innocence, a cornerstone of democratic legal frameworks codified in Article 48 of the Charter of Fundamental Rights of the European Union and Article 6(2) of the European Convention on Human Rights (ECHR). When AI systems penalise individuals for behaviour deemed statistically indicative of risk, without any formal accusation or legal process, they effectively invert the burden of proof, bypassing essential procedural safeguards (Mantelero, 2018).

Third, the implementation of social scoring exerts a chilling effect on freedoms of expression, association, and movement. Individuals may self-censor, avoid legitimate protest, or withdraw from social engagement due to fear of algorithmic penalties or reputational harm. This contributes to a climate of surveillance and conformity that is antithetical to pluralism, civic participation, and the preservation of human dignity (Zuboff, 2019).

3.3.3. The Ethical Challenge of Quantifying Human Worth

At a deeper level, social scoring systems represent an attempt to numerically quantify moral or civic worth numerically, effectively creating hierarchies of citizenship based on data traces. This datafication of reputation is both ethically troubling and epistemologically flawed. As Eubanks (2018) argues, such systems convert poverty into a predictive signal of deviance, entrenching inequality under the guise of algorithmic objectivity.

From a philosophical standpoint, these systems violate the Kantian imperative to treat individuals as ends in themselves, not as means to predictive generalizations (Kant, 1785/1996). They also breach the virtue ethics principle of moral discernment, where context, intent, and human judgment must prevail over statistical abstractions (Moor, 2006).

By prohibiting such practices, Article 5 of the EU AI Act asserts a strong moral and legal stance: algorithmic scoring of human beings by the state is incompatible with the principles of democracy, dignity, and justice.

3.4. Biometric Surveillance in Public Spaces

3.4.1. The Rise of Real-Time Biometric Surveillance

One of the most controversial AI applications addressed in Article 5(1)(d) of the EU Artificial Intelligence Act is the use of real-time remote biometric identification (RBI) systems in publicly accessible spaces for law enforcement purposes. These systems—especially facial recognition technologies (FRTs)—have gained traction globally for use in policing, border control, and public safety initiatives. They function by scanning the faces of passersby, comparing them to databases, and flagging potential matches in real time.

While some narrowly defined exceptions are allowed under the AI Act, such as for the search of missing children or prevention of terrorist threats, the default position is prohibition due to the severe and disproportionate risks posed to fundamental rights (European Commission, 2021). This cautious approach underscores the ethical and legal recognition that unregulated biometric surveillance transforms public spaces into zones of constant monitoring, deeply affecting privacy, autonomy, and democratic participation.

3.4.2. Tensions with Privacy and Human Autonomy

The right to privacy is at the heart of the opposition to biometric surveillance. Article 7 of the Charter of Fundamental Rights of the European Union guarantees the right to respect for private and family life. At the same time, Article 8 provides the right to the protection of personal data (European Union, 2012). Facial recognition technologies, especially in public areas, collect highly sensitive biometric data without the knowledge or consent of individuals. As explained by

Mantelero (2018), biometric data is not merely personal—it is immutable and identity-defining, making its misuse especially harmful and irreversible.

Moreover, biometric surveillance systems often operate with limited transparency and poor accuracy across demographic groups, exacerbating risks for racial, gender, and age-based discrimination (Buolamwini and Gebru, 2018). Studies have shown that commercial facial recognition systems exhibit higher error rates for women and people with darker skin tones, raising serious concerns about algorithmic bias and disproportionate targeting (Raji and Buolamwini, 2019). This intersection of surveillance and discrimination amounts to a violation of the right to non-discrimination under Article 21 of the Charter, as well as a breach of human dignity.

3.4.3. Threats to Freedom of Assembly and Expression

Biometric surveillance in public spaces also presents a chilling effect on freedom of assembly and expression, especially when deployed during protests, demonstrations, or politically sensitive events. The European Court of Human Rights has ruled that surveillance measures can constitute a violation of Articles 10 and 11 of the European Convention on Human Rights (ECHR), which protect freedom of expression and association, if they are disproportionate or lack sufficient legal safeguards (ECtHR, 2008).

The panoptic environment created by real-time biometric tracking undermines individuals' willingness to freely participate in democratic life, particularly when they fear being flagged, recorded, or profiled by the state (Zuboff, 2019). This extends beyond actual abuse to the perception of constant visibility, which has been shown to erode trust in public institutions and reduce civic engagement (Feldstein, 2021).

3.4.4. Surveillance Overreach and Rule of Law Concerns

Perhaps most importantly, real-time biometric surveillance shifts the balance of power decisively in favour of the state, introducing the potential for unchecked authoritarianism and surveillance overreach. The deployment of such technologies without rigorous legal safeguards, such as judicial authorisation, data minimisation, and independent oversight, violates the rule of law principle foundational to EU governance (Wright and Kreissl, 2015). Moreover, mission creep—the tendency for surveillance tools to be repurposed for broader and less justified uses—further increases the risk of abuse.

By preemptively banning these systems under most circumstances, the EU AI Act sets a clear normative threshold, asserting that public safety cannot come at the cost of fundamental freedoms.

4. Historical Misuse Cases

The prohibitions outlined in Article 5 of the EU Artificial Intelligence Act are not only anticipatory but also reactive, grounded in well-documented historical abuses of data-driven and algorithmic systems. These case studies, drawn from domains such as political manipulation, law enforcement, and public governance, illustrate the urgent need for binding legal boundaries around high-risk AI practices. They provide empirical justification for the legal classification of "unacceptable risk" AI systems, particularly those that compromise privacy, autonomy, non-discrimination, and democratic participation.

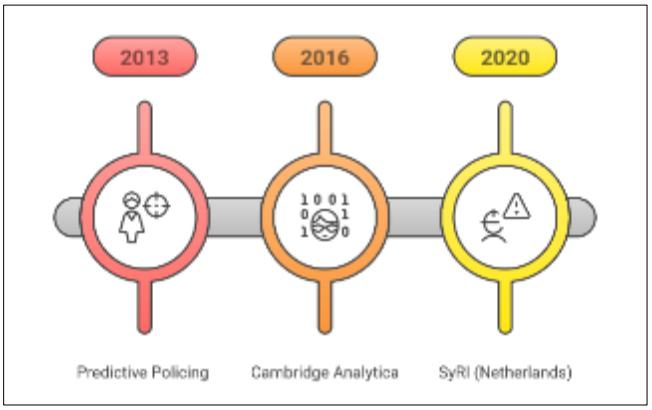

Figure 2 Timeline of Key Historical AI Misuse Cases and Their Human Rights Implications

4.1. Cambridge Analytica and Psychographic Profiling

One of the most infamous examples of AI-adjacent manipulation is the Cambridge Analytica scandal, which came to light in 2018. The political consultancy firm harvested the personal data of up to 87 million Facebook users without consent. It used it to build psychographic profiles capable of predicting personality traits, emotional vulnerabilities, and political beliefs (Isaak and Hanna, 2018). These insights were then used to micro-target users with persuasive political content, particularly during the Brexit referendum and the 2016 U.S. presidential election.

While the techniques did not rely solely on AI, the infrastructure leveraged machine learning algorithms to profile and influence individuals below the threshold of their conscious awareness. This case exemplifies the dangers of subliminal manipulation, a practice now explicitly banned under Article 5(1)(a) of the AI Act. As Zuboff (2019) explains, such techniques convert personal experience into behavioural data that can be used for surveillance-based behavioural modification, eroding mental autonomy and democratic integrity.

4.2. Predictive Policing and Racial Bias in the United States

The deployment of predictive policing algorithms in various U.S. cities over the past decade has led to widespread concerns about algorithmic discrimination and systemic bias. Tools such as PredPol have been utilised to forecast crime hotspots and recommend where police should patrol, based on historical arrest data. However, numerous studies have shown that such systems often amplify existing racial and socio-economic biases, disproportionately targeting Black and Latino communities while failing to account for unequal policing practices in the historical data (Lum and Isaac, 2016).

These systems undermine the principle of equality before the law and violate the presumption of innocence, as individuals are indirectly penalised based on group affiliations and location rather than specific actions. In some cases, individuals were placed under surveillance or flagged for increased scrutiny without evidence of wrongdoing, echoing the concerns addressed in the AI Act's ban on exploitative AI and social scoring mechanisms (Article 5(1)(b–c)). As Eubanks (2018) argues, predictive tools often function as digital poorhouses, intensifying control over already marginalized populations.

4.2.1. Informing the EU AI Act's Prohibitions

These historical cases have collectively shaped the EU's decision to classify certain AI systems as categorically impermissible. They demonstrate that data-driven manipulation, discrimination, and surveillance are not hypothetical risks, but real harms already realized in diverse political and technological contexts. The prohibitions in Article 5 reflect a synthesis of empirical evidence and ethical reasoning, emphasising that some AI uses are inherently incompatible with human rights protections, regardless of context or oversight.

Rather than relying solely on risk mitigation strategies, the AI Act draws a moral boundary, affirming that respect for privacy, autonomy, and equality must precede innovation. These precedents serve not only as cautionary tales but as guiding lights for shaping responsible AI futures.

5. Legal Precedents

The formulation of Article 5 of the EU Artificial Intelligence Act (AI Act), which identifies a select category of AI applications as posing an "unacceptable risk," did not emerge in a legal vacuum. Instead, it is deeply informed by jurisprudential precedents established by the Court of Justice of the European Union (CJEU) and the European Court of Human Rights (ECtHR). These courts have consistently upheld fundamental rights related to privacy, data protection, and non-discrimination, offering interpretive guidance that shaped the moral and legal rationale for the Act's most stringent prohibitions.

5.1. Digital Rights Ireland v. Minister for Communications (CJEU, 2014)

One of the most influential rulings shaping the AI Act's trajectory is the Digital Rights Ireland case, in which the CJEU invalidated the EU Data Retention Directive (2006/24/EC) on the grounds that it violated Articles 7 and 8 of the Charter of Fundamental Rights of the European Union—the rights to private life and protection of personal data (CJEU, 2014). The Court found that the indiscriminate and generalised retention of communication metadata without sufficient safeguards amounted to a serious interference with fundamental rights, even if the data were not directly sensitive.

This ruling is significant because it established that mass surveillance without strict proportionality and necessity requirements is incompatible with EU law. The reasoning in this case is echoed in Article 5(1)(d) of the AI Act, which

prohibits real-time remote biometric identification systems in public spaces, except in narrowly defined circumstances. As Mantelero (2018) argues, the CJEU's decision affirmed that digital tools capable of omnipresent surveillance must be constrained by constitutional principles, particularly when deployed by state authorities.

5.2. S. and Marper v. United Kingdom (ECtHR, 2008)

In the landmark S. and Marper v. UK decision, the ECtHR ruled that the indefinite retention of DNA samples, fingerprints, and biometric data of individuals not convicted of a crime violated Article 8 of the European Convention on Human Rights (ECHR), which guarantees the right to respect for private life (ECtHR, 2008). The Court emphasized that biometric data are uniquely sensitive, containing vast amounts of personal information, and that their retention without justification disproportionately encroaches on individual privacy.

This judgment directly supports the AI Act's restriction on biometric surveillance, especially facial recognition in public spaces. The ECtHR's ruling underscores the principle that the mere capability of a system to intrude into private life, not just its actual misuse, is grounds for legal prohibition. In this light, the precautionary logic of Article 5 aligns with established European legal doctrine that pre-emptively restricts technologies whose core function involves the erosion of fundamental rights (Wright and Kreissl, 2015).

5.3. Broader Legal Influence on Article 5 Drafting

Beyond these two landmark cases, the general body of European constitutional jurisprudence has consistently favoured the prioritisation of human dignity and autonomy over technological expediency. For instance, the Google Spain v. AEPD and Mario Costeja González (2014) ruling introduced the concept of the "right to be forgotten," reinforcing the notion that data subjects retain control over their digital representations (CJEU, 2014). This principle is echoed in Article 5's emphasis on preventing manipulation, exploitation, and social stratification via AI.

The AI Act's drafters appear to have internalised these precedents by embedding strict legal thresholds for specific applications, choosing outright bans where prior rulings had exposed the limits of proportionality-based regulation. As Edwards (2022) observes, Article 5 reflects an evolution in EU digital law: from procedural safeguards and balancing tests toward substantive limits on what kinds of AI systems should be built and deployed at all.

6. Human Impact Analysis

While legal and ethical debates surrounding the EU Artificial Intelligence Act (AI Act) provide a regulatory framework, the lived human impacts of the prohibited practices listed in Article 5 are crucial to understanding their severity. The harms associated with subliminal manipulation, exploitation of vulnerable groups, social scoring, and biometric surveillance extend beyond abstract rights violations, manifesting in psychological trauma, social exclusion, civic disengagement, and structural inequality. This section explores four key areas where these AI applications compromise human dignity and autonomy.

6.1. Psychological Harm: Behavioural Manipulation and the Hypernudge

The concept of "nudging"—subtle interventions designed to influence decision-making—has long been used in policy design. However, when amplified through AI technologies that continuously monitor and predict individual behaviour, nudging can morph into what Karen Yeung (2017) terms the "hypernudge". Hypernudging operates via real-time feedback loops, adaptive personalisation, and data-driven profiling, often without the user's awareness. These systems are capable of reshaping preferences and behaviours in ways that bypass rational deliberation and consent.

Subliminal manipulation, as prohibited in Article 5(1)(a), reflects the growing concern that such techniques compromise mental autonomy—the freedom to think and act without covert coercion. The psychological harm is especially acute among children and cognitively vulnerable individuals, who lack the cognitive defences to resist behavioural steering. As Susser, Roessler, and Nissenbaum (2019) argue, this constitutes an attack on the "freedom of thought", a foundational element of dignity and democratic self-determination.

6.2. Social Harm: Discrimination, Exclusion, and Life Opportunity Reduction

The deployment of AI in welfare, education, and criminal justice, particularly in the form of predictive analytics or risk scoring, has been shown to exacerbate existing social inequalities. When individuals are flagged as "high risk" based on opaque or biased data, they may face denial of services, intensified scrutiny, or exclusion from opportunities (Eubanks, 2018). These harms disproportionately affect already marginalized communities, including racial minorities, low-income individuals, and persons with disabilities.

Systems that exploit vulnerabilities or socially score individuals, as banned under Articles 5(1)(b) and 5(1)(c), introduce a feedback loop of disadvantage. For example, individuals labelled untrustworthy by AI-driven credit or policing systems are likely to experience compounded social penalties, reducing access to housing, employment, and education (Wachter et al., 2021). The result is not only discrimination, but the automated reproduction of social stratification.

6.3. Chilling Effect: Surveillance and the Erosion of Expression and Protest

Real-time biometric surveillance, especially in public spaces, generates a chilling effect on democratic participation. As individuals become aware of constant monitoring, they may alter their behaviour, not only online but also in physical environments. Research has shown that people under surveillance are less likely to attend protests, express dissent, or associate with stigmatized groups for fear of being recorded, profiled, or flagged (Feldstein, 2021).

This fear-driven self-censorship directly undermines the freedom of expression and assembly, which are protected under both the EU Charter and the European Convention on Human Rights. The panoptic conditions created by facial recognition systems represent not just an invasion of privacy but a deterrent to civic courage, disproportionately affecting activists, journalists, and vulnerable populations (Zuboff, 2019).

6.4. Digital Divide: Unequal Exposure to Algorithmic Harm

The risks posed by prohibited AI systems are not distributed evenly. Individuals lacking digital literacy, technological access, or institutional support are significantly more vulnerable to uninformed consent, algorithmic misclassification, and systemic exclusion. For example, job seekers unfamiliar with automated hiring platforms may not understand how their data is being analysed or rejected, while low-income users of public services may be unaware of how algorithms assess their eligibility or flag them for fraud investigations (Eubanks, 2018).

This digital divide reinforces broader socio-economic disparities. The prohibited practices under Article 5 tend to compound preexisting inequities, turning technical systems into mechanisms of deepened structural injustice (Lepri et al., 2018). Addressing these harms requires not only banning abusive systems but ensuring inclusive digital governance, public education, and accessible redress mechanisms for those impacted.

7. Risk Assessment of Current AI Systems

Contemporary artificial intelligence systems, especially those integrated into social, economic, and legal domains, present a spectrum of ethical risks ranging from data misuse to covert behavioural manipulation. The EU Artificial Intelligence Act introduces a tiered regulatory approach that classifies these risks. Among these, systems deemed to pose an "unacceptable risk" under Article 5 are explicitly prohibited due to their capacity to undermine fundamental human rights such as autonomy, privacy, dignity, and equality (European Commission, 2024).

7.1. Emotion Inference in Educational and Workplace Settings

AI systems that attempt to infer emotions from facial expressions or physiological signals are increasingly being deployed in both schools and professional environments. These systems claim to detect engagement, stress, or motivation; however, the scientific basis for such inferences is widely contested. Emotion is not universally expressed or recognized, and such systems often exhibit cultural bias and low reliability (Barrett et al., 2019). Article 5(1)(f) prohibits the use of emotion recognition systems in education and the workplace, unless justified by safety or medical necessity, due to the potential for unjust profiling, psychological harm, and infringements on mental privacy.

7.2. Profiling and Predictive Policing

Predictive profiling systems are used to assess the likelihood of individuals engaging in future behavior based on statistical correlations and inferred traits. While sometimes framed as tools to enhance public safety, these systems risk penalizing individuals based on probabilistic assumptions rather than concrete actions. Such uses are deeply problematic when they draw conclusions based solely on personality traits or behavioral profiling. Article 5(1)(d) bans systems that attempt to assess criminal risk without objective, verifiable evidence, as they compromise the presumption of innocence and disproportionately affect marginalized populations (Mantelero, 2018).

7.3. Exploitation of Vulnerable Groups

Some AI applications are designed to influence or direct the decisions of specific populations, such as children, the elderly, or economically disadvantaged individuals. These systems may leverage user vulnerabilities—intentionally or otherwise—to encourage certain behaviors, such as extended screen time, overconsumption, or over-reliance on

automation. Article 5(1)(b) prohibits systems that exploit such vulnerabilities when they result in behavioral distortion or significant harm. These systems raise deep ethical concerns about manipulation and coercion, especially in contexts where informed consent is difficult to ensure (Floridi et al., 2018).

7.4. Biometric Surveillance in Public Spaces

AI systems capable of identifying individuals in real time using biometric data—such as facial recognition—have become increasingly common in public safety, transportation, and urban monitoring. Despite claims of utility, these systems pose serious threats to anonymity in public life and open the door to mass surveillance. Article 5(1)(h) prohibits real-time remote biometric identification systems in public spaces for law enforcement purposes, except under narrowly defined exceptions with strong safeguards, including prior judicial authorization. Without strict limitations, such technologies risk normalizing a surveillance culture that is antithetical to democratic freedoms (Zuboff, 2019).

7.5. Risk Amplification Through Data Aggregation

Some AI systems aggregate vast datasets from diverse sources, often without transparent mechanisms or adequate consent. When used to evaluate individuals or predict behaviors, these systems can generate risk scores that influence access to services, opportunities, or legal outcomes. Although not always explicitly prohibited, these systems may veer into social scoring or exploitative practices if they involve non-contextual data use or discriminatory logic. Their risk profile must be assessed in relation to Article 5(1)(c), which bans unjustified or disproportionate social scoring by public authorities (Wachter et al., 2021).

Current AI systems often operate at the edge of what is considered acceptable in democratic societies. Article 5 of the EU AI Act provides a critical framework to identify and prohibit the most egregious uses. However, the evolving nature of AI requires continual reassessment of risk, informed by interdisciplinary research, human rights jurisprudence, and empirical evidence.

8. Discussion

The prohibited practices outlined in Article 5 of the EU AI Act represent more than a technical categorization of high-risk systems—they reflect a broader commitment to safeguarding fundamental human rights in the face of rapid technological development. Each prohibited application threatens fundamental rights, including privacy, autonomy, equality, freedom of expression, and human dignity. Subliminal manipulation, for example, undermines the principle of informed consent by influencing behaviour outside an individual's conscious awareness, violating the autonomy protected by Article 1 and Article 8 of the Charter of Fundamental Rights of the European Union (CFR). Similarly, the exploitation of vulnerable individuals, such as children, the elderly, or socio-economically disadvantaged groups, raises concerns over fairness and equal protection under Article 21 of the CFR. These practices not only distort agency but also risk reinforcing systemic inequalities, especially when embedded in opaque algorithmic structures (Floridi et al., 2018; Wachter et al., 2021).

Social scoring presents an even broader challenge to democratic values, threatening the presumption of innocence and creating a chilling effect on lawful behaviour. By assigning behavioural scores to individuals, these systems can result in discrimination and social exclusion without due process, directly contravening Article 48 of the CFR and Article 6(2) of the European Convention on Human Rights (Mantelero, 2018). Biometric surveillance in public spaces, including real-time facial recognition by law enforcement, infringes on privacy and freedom of assembly, raising particular concerns about mass surveillance and the normalisation of panoptic oversight. Other banned practices—such as untargeted scraping of facial images or emotion recognition in sensitive settings like schools or workplaces—undermine the integrity of identity, personal space, and scientific legitimacy, often operating on unverified or pseudo-scientific claims.

The ethical and legal frameworks that justify the prohibition of these systems are grounded in both binding rights instruments and normative principles. Legally, the EU's CFR and the ECHR provide the backbone for rights-based regulation, particularly emphasising dignity, justice, privacy, and non-discrimination. Ethically, global declarations such as the UNESCO Recommendation on the Ethics of AI (2021) and the OECD AI Principles (2019) emphasize the need for explainability, accountability, and respect for human autonomy. These frameworks converge on the idea that AI should serve human flourishing, rather than instrumentalising, profiling, or manipulating individuals. The "ethics by design" approach advocated in the literature demands that AI development begins with, and remains accountable to, these core values (Floridi et al., 2018).

Real-world cases demonstrate how the absence of explicit prohibitions has already led to significant harm. Predictive policing systems have disproportionately targeted racialised communities, reinforcing historical bias embedded in data

and institutional practices (Angwin et al., 2016). Affective computing systems marketed for use in employment or educational evaluations have drawn criticism for attempting to infer internal states, such as motivation, trustworthiness, or engagement, using methods not supported by scientific consensus (Williams et al., 2022). Large-scale biometric data scraping for surveillance or marketing purposes has occurred in jurisdictions lacking robust data protection laws, raising concerns about consent and potential misuse. These examples demonstrate that without pre-emptive regulation, such as Article 5, AI technologies risk becoming tools of systemic injustice and control.

Despite Article 5's prohibitions, some current AI systems continue to approach or blur ethical boundaries. Real-time biometric surveillance is still deployed in public safety contexts under temporary legal exemptions or soft mandates, with insufficient oversight. Emotion recognition systems—especially those targeting facial microexpressions—are marketed to schools and employers, where users may not have the knowledge or power to refuse consent. Behavioural scoring systems are increasingly used in financial and insurance sectors, creating outcomes that resemble social scoring logic without formal state endorsement. These cases illustrate a regulatory lag between ethical awareness and policy enforcement. As Wachter et al. (2021) argue, fairness and non-discrimination cannot be reliably outsourced to algorithms without legal and institutional safeguards.

Ultimately, the risk assessment of current AI systems must consider not only technical reliability but also the broader social implications. Article 5 of the EU AI Act serves as a model for articulating clear red lines—practices that are inherently incompatible with democratic values. While enforcement challenges remain, the Act signals a shift from abstract principles to actionable legal prohibitions. This transformation is essential if AI is to be developed in a way that respects and protects the rights of all individuals, particularly the most vulnerable.

9. Conclusion

This review critically examined the prohibited practices outlined in Article 5 of the European Union Artificial Intelligence Act (EU AI Act), with a focus on their ethical justifications and potential human rights violations. The analysis revealed that each of the eight banned AI applications—ranging from subliminal manipulation to biometric surveillance—poses a substantive threat to foundational rights such as dignity, autonomy, equality, privacy, and freedom of expression. These threats are not hypothetical; they are grounded in empirical misuse cases, legal precedent, and consistent ethical reasoning.

This study aimed to bridge the gap between legal text and human rights analysis by systematically assessing why these practices have been classified as carrying an "unacceptable risk." Each prohibition reflects a more profound concern: not merely about technology itself, but about the social, legal, and ethical implications of its deployment. As such, the findings contribute to both AI governance literature and human rights scholarship by showing that these AI systems undermine not only technical safeguards, but also the moral fabric of democratic societies.

The findings confirm that AI systems capable of manipulating human behaviour, profiling vulnerable populations, or subjecting individuals to opaque scoring mechanisms erode core democratic protections. These systems circumvent informed consent, ignore contextual fairness, and often operate without meaningful oversight or recourse. More importantly, they tend to disproportionately harm the most vulnerable—those already at the margins of social, economic, or political life. The prohibitions, therefore, function as both legal constraints and normative statements, declaring that some technological paths are incompatible with the values we seek to uphold.

A key insight from this review is the growing convergence between ethical theory and human rights law. Deontological ethics, utilitarianism, and virtue ethics each offer distinct but overlapping arguments in support of these bans. Whether by highlighting the violation of autonomy, the disproportionality of harms, or the erosion of social trust, ethical frameworks reinforce the legal rationale behind the prohibitions. At the same time, jurisprudence from European courts—such as the Digital Rights Ireland and S. and Marper cases—provides concrete legal grounding for constraints on AI surveillance and profiling technologies.

This review has several practical implications. For regulators, it emphasizes the importance of moving beyond abstract ethical principles to concrete prohibitions backed by enforceable law. For developers and technology companies, it underscores the importance of engaging in human rights due diligence and avoiding the design of systems that may later be deemed unlawful. For civil society, it provides a foundation for advocacy that is grounded in both empirical evidence and moral theory. Ultimately, for international policymakers, the EU's model offers a scalable framework that could inform global AI governance efforts, particularly in jurisdictions lacking robust human rights protections.

Nonetheless, the study has limitations. It does not include quantitative metrics of harm, nor does it assess how these prohibitions are currently being enforced. Furthermore, while the review draws on EU-based legal and ethical frameworks, broader perspectives—including those from non-Western legal traditions—remain underexplored.

Future research should investigate how AI systems evolve in response to legal prohibitions and whether new forms of manipulation or discrimination emerge in more subtle or decentralised ways. There is also a need for empirical studies assessing the lived impact of prohibited AI technologies and for comparative legal research on how other regions define and regulate high-risk AI.

In conclusion, Article 5 of the EU AI Act represents a significant shift from soft ethics to complex law. It draws a moral and legal boundary around practices that compromise human dignity and democratic integrity. This review highlights that prohibiting specific AI systems is not anti-innovation—it is a declaration of the kind of society we want to build in an age of rapid technological transformation.

By clarifying the ethical and legal foundations for prohibiting high-risk AI systems under the EU AI Act, this study contributes to the responsible development of AI. It informs global policymakers, ultimately promoting a future where technological innovation aligns with human dignity and democratic values.

Compliance with ethical standards

Disclosure of conflict of interest

The author declares no conflict of interest.

Statement of Ethical Approval

This article does not contain any studies with human participants or animals performed by the author.

Funding

No external funding was received for the preparation of this manuscript.

Data Availability Statement

No datasets were generated or analysed during the current study.

References

[1] Ajunwa, I. (2020). The paradox of automation as anti-bias intervention. Cardozo Law Review, 41(5), 1671–1738. https://cardozolawreview.com/the-paradox-of-automation-as-anti-bias-intervention/

[2] Angwin, J., Larson, J., Mattu, S., and Kirchner, L. (2016). Machine bias. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[3] Binns, R. (2019). On the apparent conflict between individual and group fairness. arXiv. https://doi.org/10.48550/arXiv.1912.06883

[4] Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. In S. A. Friedler and C. Wilson (Eds.), Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency (FAT)* (Vol. 81, pp. 149–159). PMLR. https://doi.org/10.48550/arXiv.1712.03586

[5] Buolamwini, J., and Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In S. A. Friedler and C. Wilson (Eds.), Proceedings of the 1st Conference on Fairness, Accountability, and Transparency (Vol. 81, pp. 77–91). PMLR. https://proceedings.mlr.press/v81/buolamwini18a.html

[6] Butcher, J., and Beridze, I. (2019). What is the state of artificial intelligence governance globally? The RUSI Journal, 164(5–6), 88–96. https://doi.org/10.1080/03071847.2019.1694260

[7] Cihon, P., Maas, M. M., and Kemp, L. (2020). Should artificial intelligence governance be centralised? Design lessons from history. Proceedings of the 2020 AAAI/ACM Conference on AI, Ethics, and Society (AIES '20), 7–8 February 2020, New York, NY. Association for Computing Machinery. https://doi.org/10.1145/3375627.3375857

[8] Court of Justice of the European Union. (2014). Digital Rights Ireland Ltd v. Minister for Communications, Marine and Natural Resources and Others and Kärntner Landesregierung and Others (Joined Cases C-293/12 and C-594/12). ECLI:EU:C:2014:238. https://curia.europa.eu/juris/liste.jsf?num=C-293/12

[9] Crawford, K. (2021). Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press.

[10] European Court of Human Rights. (2008). S. and Marper v. the United Kingdom (Applications Nos. 30562/04 and 30566/04). ECLI:CE:ECHR:2008:1204JUD003056204. https://hudoc.echr.coe.int/fre?i=001-90051

[11] Edwards, L. (2022). The EU AI Act: A summary of its significance and scope. Ada Lovelace Institute. https://www.adalovelaceinstitute.org/wp-content/uploads/2022/04/Expert-explainer-The-EU-AI-Act-11-April-2022.pdf

[12] Eubanks, V. (2018). Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press.

[13] European Commission. (2021). Proposal for a regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021) 206 final). EUR-Lex. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

[14] European Commission. (2024). Artificial Intelligence Act – Article 5. Retrieved from https://artificialintelligenceact.eu/article/5/

[15] European Union. (2012, October 26). Charter of Fundamental Rights of the European Union (2012/C 326/02). Official Journal of the European Union, C 326, 391–407. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:C2012/326/02

[16] Feldstein, S. (2021). The Rise of Digital Repression: How Technology is Reshaping Power, Politics, and Resistance. Oxford University Press.

[17] Floridi, L. (2013). The Ethics of Information. Oxford University Press.

[18] Floridi, L., and Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review, 1(1). https://doi.org/10.1162/99608f92.8cd550d1

[19] Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... and Schafer, B. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

[20] Isaak, J., and Hanna, M. J. (2018). User data privacy: Facebook, Cambridge Analytica, and privacy protection. Computer, 51(8), 56–59. https://doi.org/10.1109/MC.2018.3191268

[21] Kant, I. (1996). Groundwork of the Metaphysics of Morals (M. Gregor, Trans.). Cambridge University Press. (Original work published 1785)

[22] Lepri, B., Oliver, N., Letouzé, E., Pentland, A., and Vinck, P. (2018). Fair, transparent, and accountable algorithmic decision-making processes. Philosophy and Technology, 31(4), 611–627. https://doi.org/10.1007/s13347-017-0279-x

[23] Lum, K., and Isaac, W. (2016). To predict and serve? Significance, 13(5), 14–19. https://doi.org/10.1111/j.1740-9713.2016.00960.x

[24] Madaio, M. A., Stark, L., Wortman Vaughan, J., and Wallach, H. (2020). Co-designing checklists to understand organizational challenges and opportunities around fairness in AI. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1–14). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376445

[25] Mantelero, A. (2018). Report on artificial intelligence: Artificial intelligence and data protection—Challenges and possible remedies. Council of Europe. https://rm.coe.int/report-on-artificial-intelligence-artificial-intelligence-and-data-pro/16808e6012

[26] Mittelstadt, B. (2017). Ethics of the health-related internet of things: A narrative review. Ethics and Information Technology, 19(3), 157–175. https://doi.org/10.1007/s10676-017-9426-4

[27] Moor, J. H. (2006). The nature, importance, and difficulty of machine ethics. IEEE Intelligent Systems, 21(4), 18–21. https://doi.org/10.1109/MIS.2006.80

[28] Raji, I. D., and Buolamwini, J. (2019). Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, 429–435. https://doi.org/10.1145/3306618.3314244

[29] Ruggie, J. G. (2011). Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework. United Nations Human Rights Council.

[30] United Nations Human Rights Council. (2011). Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework. Office of the High Commissioner for Human Rights. https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf

[31] Selwyn, N. (2020). Should Robots Replace Teachers? AI and the Future of Education. Polity Press.

[32] Susser, D., Roessler, B., and Nissenbaum, H. F. (2019). Online manipulation: Hidden influences in a digital world. Georgetown Law Technology Review, 4(1), 1–45. https://doi.org/10.2139/ssrn.3306006

[33] Taddeo, M., and Floridi, L. (2018). How AI can be a force for good. Science, 361(6404), 751–752. https://doi.org/10.1126/science.aat5991

[34] Tufekci, Z. (2015). Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency. Colorado Technology Law Journal, 13(2), 203–218. https://ctlj.colorado.edu/wp-content/uploads/2015/08/Tufekci-final.pdfctlj-dev.cu.law+10

[35] United Nations Educational, Scientific and Cultural Organization. (2021). Recommendation on the ethics of artificial intelligence. https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence

[36] Allen, R., and Masters, D. (2020, March 30). SyRI: Think twice before risk profiling. AI Law Consultancy. https://ai-lawhub.com/spring-2020-syri-judgment/

[37] Veale, M., and Borgesius, F. Z. (2021). Demystifying the Draft EU Artificial Intelligence Act. Computer Law Review International, 22(4), 97–112. https://doi.org/10.9785/cri-2021-220402

[38] Wachter, S., and Mittelstadt, B. (2019). A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI. Columbia Business Law Review, 2019(2), 494–620. https://doi.org/10.2139/ssrn.3248829

[39] Wachter, S., Mittelstadt, B., and Russell, C. (2021). Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. Computer Law and Security Review, 41, 105567. https://doi.org/10.1016/j.clsr.2021.105567

[40] Whittlestone, J., Nyrup, R., Alexandrova, A., and Cave, S. (2019). The role and limits of principles in AI ethics: Towards a focus on tensions. In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (pp. 195–200). Association for Computing Machinery. https://doi.org/10.1145/3306618.3314289

[41] Wright, D., and Kreissl, R. (Eds.). (2014). Surveillance in Europe. Routledge.

[42] Wright, D., and Raab, C. D. (2012). Constructing a surveillance impact assessment. Computer Law and Security Review, 28(6), 613–626. https://doi.org/10.1016/j.clsr.2012.09.003

[43] Yeung, K. (2017). 'Hypernudge': Big data as a mode of regulation by design. Information, Communication and Society, 20(1), 118–136. https://doi.org/10.1080/1369118X.2016.1186713

[44] Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. PublicAffairs.