17th Cornell Law School Graduate Conference

March 13-14, 2025

Dear all,

Thank you for the opportunity to discuss this paper with you. As you will see, this paper is at a preliminary stage and aims and is a tentative to provide structure to my evolving research ideas.
It is part of a broader research project I am conducting on AI regulation and the misuse of AI by governments. In this context, the paper serves as an initial attempt to frame key issues and develop arguments that I hope to refine through further research and discussion.

Please note that some parts need to be expanded, and the conclusion likely needs to be deepened. In this regard, I would greatly welcome your suggestions on additional bibliography that could be useful, particularly from a philosophy of law perspective. I would also appreciate any recommendations for additional examples that could serve as compelling case studies to illustrate the issues at stake.
Furthermore, while my current analysis is primarily focused on the European Union's AI Act (which I consider offering a sufficiently robust framework for advancing my claims) I would be very interested in hearing your opinions on whether incorporating perspectives from other jurisdictions could enrich the analysis.

Thank you again for your time and I and look forward to our discussion.

Best,
Federica Fedorczyk

NOT A RATIONAL CHOICE: HOW AI (REGULATION) IS ERODING OUR AUTONOMY

Federica Fedorczyk*

Abstract

Legal systems as we know them are about to change with the advent of AI, and public regulation worldwide may not be ready.

In my paper I critically analyze the way AI has been regulated until now, describing how – although the serious risks of certain AI applications are clear – governments and legislators across the globe have deliberately chosen to accept these risks at the expense of fundamental rights. They profess a commitment to human rights, while simultaneously justifying their erosion or denial.

I claim that this tendency is not uncommon and permeates not only the relationship between AI and regulators but also between AI and individuals. At an individual level, we are aware of the serious risks associated with certain AI applications, but we often forgo self-protection for the lure of immediate benefits. In doing so, we might even compromise our own values - much like governments that might choose to sideline fundamental rights to prioritize other goals.

The appeal of AI's immediate and long-term benefits – such as enhanced performance, increased efficiency and safety, along with advancements in healthcare and production – gradually reduces our willingness to make sacrifices to uphold our other core values. Though we value our privacy, we freely share personal information with chatbots; though we value our intelligence and critical thinking, we increasingly rely on tools like ChatGPT. Our reasons may vary, but they converge under the same umbrella: performance benefits.

This mirrors a tendency also seen among regulators: while professing to prioritize democracy and individual freedoms, they allow and even promote AI systems that pose significant threats to these very principles. The rationale is often rooted in securing different benefits, such as increased public safety and control, that supposedly contribute to overall societal performance.

In the paper, I critically examine the following persistent misconception: that both individuals and governments make rational and balanced choices when deciding to use AI, weighing its benefits against its risks.

I argue the opposite, making the following hypothesis: the rational choice is progressively becoming not possible. The very existence of AI - with its significant advantages - subtly erodes our ability to truly choose, as our dependence on these tools increasingly narrows the scope of viable alternatives.

Building on this hypothesis, the paper aims to explore whether AI's transformation from an optional resource into a perceived necessity – often an assumed and inescapable condition of human life – truly results in a diminished capacity for individual and government autonomy.

In examining this shift, I focus on the possibility that, as individuals increasingly delegate their autonomy, critical thinking and decision-making to AI, a parallel erosion is occurring at the level of regulators. The latter poses an even greater risk, as it weakens or even disintegrates the essential role of regulators as the protector of its citizens and society's core values, reshaping the traditional relationship of trust between governors and the governed. As a result, trust deteriorates both vertically, as individuals lose faith in their governments, and horizontally, as they doubt their own and their peers' capacity for self-determination, critical thinking, and autonomy.

Keywords: AI regulation; fundamental rights; autonomy erosion; trust and dependence.

Conference theme: Philosophy of Law

* Federica Fedorczyk is a Postdoctoral Emile Noël Fellow at NYU, where she is also an Affiliate at the Information Law Institute (ILI), and a Postdoctoral Research Fellow at Sant'Anna School of Advance Studies. Starting in June 2025, she will join the University of Oxford as an Early Career Research Fellow at the Institute for Ethics in AI.

1. Introduction

This paper intends to examine how legislators and regulators approach AI regulation without adequately considering the risks associated with its deployment, mirroring patterns observed in individual decision-making. It does not aim to propose a regulatory strategy or to suggest policy recommendations to address the increasingly pervasive challenges posed by AI applications in various aspects of daily life.

The critical analysis of AI regulation will focus exclusively on the European Union's Artificial Intelligence Act (AIA), which serves as a case study. This choice is primarily motivated by the AIA's global reputation as one of the most protective regulatory frameworks for AI in terms of safeguarding human rights and fundamental freedoms, while allegedly hindering innovation by imposing excessively stringent requirements on AI developers and deployers (Bradford, 2023). However, while this paper is limited to the examination of the AIA, I believe the arguments presented remain broadly applicable to other AI regulations worldwide, with necessary contextual adjustments.

This analysis will highlight specific provisions within the AIA that reveal how the protection of democracy and individual freedoms has been compromised in favor of other priorities, such as economic growth, public safety, and state control, objectives often justified by their perceived contributions to overall societal well-being. Through this examination, I aim to draw a parallel between the ways individuals interact with AI and how regulators approach its governance. I argue that the immediate and long-term benefits of AI – such as enhanced performance, efficiency, safety, and advancements in healthcare and production – gradually diminish the willingness of both individuals and regulators to make sacrifices necessary to uphold other core values.

To illustrate this dynamic, consider the following examples that will be further analyzed in the following sections: despite valuing privacy, we willingly share personal information with companies or chatbots (Cofone, 2023); despite valuing intelligence and critical thinking, we increasingly rely on AI tools like ChatGPT. While the reasons for these choices may vary, they ultimately converge under a common rationale: performance benefits in terms of time savings and efficiency gains. I argue that this tendency extends to regulators as well, with the AIA serving as a prime example.

Based on this preliminary analysis, I critically examine the following persistent misconception: that both individuals and governments make rational and balanced choices when deciding to use AI, weighing its benefits against its risks. I argue the opposite, making the following hypothesis: the choice is progressively becoming not possible. The very

existence of AI - with its significant advantages - subtly erodes our ability to truly choose, as our dependence on these tools increasingly narrows the scope of viable alternatives.
We are witnessing AI's transformation from an optional resource into a perceived necessity — an assumed and inescapable condition of modern life. In exploring this shift, the final section of this paper examines a parallel phenomenon: as individuals delegate increasing aspects of their autonomy, critical thinking, and decision-making to AI, a similar erosion occurs at the regulatory level. This development poses an even greater risk, as it weakens or even disintegrates the essential role of regulators as the protector of its citizens and society's core values, reshaping the traditional relationship of trust between governors and the governed. As a result, trust deteriorates both vertically, as individuals lose faith in their governments, and horizontally, as they doubt their own and their peers' capacity for self-determination, critical thinking, and autonomy.

The paper is structured as follows. In the first section, after a brief overview of the AIA, I will identify and analyze specific provisions that illustrate how legislators have prioritized public safety, state control, and economic benefits for AI developers over the protection of fundamental rights and freedoms.
In the second section, I will examine examples of individual behaviors that demonstrate an overreliance on AI at the expense of core values, such as autonomy and self-determination.
In the third section, I will critically engage with the view that both individuals and governments make rational and balanced choices when using or regulating AI. I argue that AI's growing indispensability and appeal are leading both regulators and individuals to compromise their freedoms, security, critical thinking, and commitment to fundamental rights in pursuit of personal or collective benefits.
Finally, I will reflect on the long-term implications of this shift, exploring the risks of addiction and manipulation in a scenario where regulators fail to impose meaningful limits and safeguards on AI development and deployment.

2. AI regulation: why what has been done until now is not enough.

On July 12, 2024, the European Union reached a major milestone by approving the AIA, the world's first comprehensive framework for regulating artificial intelligence.[1] As a

[1] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act).

horizontal piece of legislation, the AIA establishes broad regulatory principles applicable across various sectors and AI applications, seeking to balance the technology's potential risks and benefits. Among its key objectives is ensuring that AI systems (AIS) introduced, deployed, and used within the Union comply with existing laws, uphold fundamental rights, and prioritize safety. At its core, the AIA envisions AI as a human-centric technology designed to enhance human well-being (Commission, 2021). Its overarching goals include improving the internal market, fostering innovation, and safeguarding health, safety, democracy, and the environment from the potential harms associated with AI systems. However, the legal basis of the regulation is Article 114 TFEU, which explains why its primary objective, as stated in the first recital, is the creation of an internal market for the free circulation of AI, placing the protection of human rights somewhat in the background. As Smuha notes, this approach is not uncommon in EU legislation, as the EU lacks a general legal basis to regulate fundamental rights, the rule of law, and democracy, often necessitating reliance on Article 114 TFEU (Smuha, 2024).

Framed as a safety-oriented piece of legislation, the AIA addresses the risk associated with the use of AI in certain contexts using a risk-based approach, which means that the content and intensity of the rules are tailored to the risks that AIS may create (Kaminski, 2023).

The idea behind a risk-based approach can be illustrated visually by a risk pyramid (Edwards, 2021): at the top are the applications that are prohibited (although there is a wide range of exceptions) because they pose an unacceptable risk (Art. 5 AIA); then there is the core of high-risk systems that, although risky, enable important functions and are therefore permitted under certain conditions (Art. 6 *et seq*. AIA); then there are limited-risk systems that must comply with the minimal transparency requirements set forth in Art. 52 of the AIA; and at the bottom, there are minimal or no risk systems that are not subject to any specific regulatory obligations under the AIA.

Although the AIA has been promoted and presented as a regulation focused on protecting fundamental rights and has been the product of numerous human-centric soft law instruments that preceded it – including the Ethics Guidelines for Trustworthy AI developed by the independent AI High Level Expert Group (HLEG, 2019) – several critical issues remain in the regulation of certain AI applications used in the law enforcement context (Díaz-Rodríguez et al., 2023)

In the final version of the AIA we can observe that tendency to accept significant risks and compromises of our fundamental values for the sake of immediate benefits. Indeed, the final version of the Act, now law, differs significantly from the version approved by the

European Parliament during the legislative process, which was more oriented towards guaranteeing greater protection of fundamental rights. Instead, the AIA reflects ultimately the Council's position more strongly than that of the Parliament. Many of the key amendments proposed in the Parliament's version of the AIA underwent further modifications during the trialogue process between the Parliament, the Council, and the Commission, and not all these amendments were incorporated into the final text, resulting in a notable reduction in the protection of fundamental rights in favor of greater power for the Member States.

To better explain the said tendency, the analysis of two examples can be useful: the regulation of remote biometric identification (RBI) systems and the regulation of manipulative AI tools.

2.1. The regulation of RBI

A useful example is the regulation of RBI. In the Parliament's version, the use of real-time RBI was banned, and the use of *ex-post* RBI was banned except in cases of severe crime and only with judicial authorization. However, in the final text of the AIA, the regulation of these applications has changed.

Real-time RBI is theoretically classified as a prohibited practice, but several exceptions still allow its use, such as preventing a genuine and foreseeable threat of a terrorist attack (Article 5, para. 1, lett. (h), point. ii). *Ex-post* RBI, instead, is classified merely as high-risk and is therefore allowed under certain conditions (Article 6, para. 2).

The decision to differentiate the regulation of real-time and *ex-post* RBI is extremely dangerous, and how the AIA addresses the use of these systems in law enforcement activity is even more concerning. Notably, even before the adoption of the AIA and during the negotiations between the various European institutions, civil society and parts of academia advocated for the ban of both real-time and retrospective RBI.(EDRi, 2023). However, these recommendations were partly ignored by the legislators, resulting in the current regulation of RBI having significant loopholes that could seriously endanger fundamental freedoms and rights.

Indeed, now the AIA introduces a sharp distinction between real-time RBI and retrospective RBI, based on their different alleged impact on fundamental rights. Although it recognizes that retrospective biometric identification systems are intrusive tools and should therefore be subject to safeguards, it classifies them simply as high-risk systems. Therefore, *ex-post* RBI can be used under certain conditions and by meeting certain requirements. Among the different critical remarks that it is possible to advance against the regulation of *ex-post* RBI, there are four aspects that are extremely problematic:

a) The fact that the authorization to use it in the framework of a criminal investigation may also be granted by a regular administrative authority chosen by Member States and not necessarily by a judicial authority or an independent administrative authority;

b) The fact that this authorization is not required to use it in the framework of a criminal investigation for the initial identification of a potential suspect based on objective and verifiable facts directly linked to the offence;

c) The possibility to use it in the law enforcement context even when there is only a genuine and foreseeable threat of a criminal offence;

d) The fact that law enforcement and migration authorities can benefit of an exemption to the requirement of the registration of the high-risks systems they use (included RBI) in a publicly accessible dataset.

a) Beginning with the first point, it is concerning that the AIA permits a regular administrative authority, rather than a judicial one or an independent administrative one, to authorize the use of retrospective RBI in criminal investigations. Clearly, this raises concerns about the impartiality of the authority called upon to grant the authorization. More specifically, the decision to leave the actual enforcement of *ex-post* RBI to the administrative authority can be extremely problematic in those countries where the repression of political dissent or, more generally, political opposition is particularly severe, such as, for instance, in Hungary.

In this regard, it should be noted that real-time RBI requires the authorization of an independent administrative authority or a judicial authority. Even this condition raises concerns, since it does not eliminate the risk that even formally independent authorities may be influenced by governments. Indeed, as has been rightly pointed out, the reliance on administrative oversight than judicial oversight may not provide sufficient checks and balances (Hacker, 2023). However, this requirement – which does not exist for *ex post* RBI – at least represents a formal obstacle to the total control of the use of RBI by a general administrative authority without further requirements.

b) The second aspect mentioned above is even more concerning. Article 26, paragraph 10 allows for the use of RBI in a criminal investigation to identify a potential suspect without authorization. The fact that the provision requires that the identification of the suspect by

RBI means should be «based on objective and verifiable facts directly linked to the offence» does not provide any assurance that these systems will not be used by police authorities on the basis of mere suspicion without the need for further control. Indeed, first, «objective and verifiable facts» is a broad concept, that does not further specify what elements should be present to transform an individual into a potential suspect. Second, there is no authority designated to control the actual existence of these (broad) conditions. Therefore, this is a huge and significant loophole that poses a severe threat to civil liberties and personal privacy and may seriously violate the presumption of innocence.

c) The most worrying provision is the one that provides the possibility to use post-remote RBI in the law enforcement context even if there is merely a genuine and foreseeable threat of a criminal offence. The wording of the provision is quite tricky and hides a strong legislative stance against not only the rights of the offender, but also, more in general, the fundamental freedoms of every individual, regardless their proven involvement in criminal acts.

In fact, referring to post-remote RBI systems, Article 26 paragraph 10 first affirms that under no circumstances should they be used for law enforcement purposes in an untargeted manner, without a link to a criminal offence or a criminal proceeding. Then, in the same sentence, it goes on to state that they can be used also if there is a link to a «genuine and present threat of a crime» or even to a «genuine and foreseeable threat of a crime».

The fact that the existence of a mere "foreseeable" threat of a crime is sufficient to use retrospective RBI opens the door to an unprecedented use of intrusive investigative tools without the existence of an actual crime that has been committed and even without the existence of a present threat of a crime.

Even the reference to a genuine and present threat of a crime (without the crime having been committed) would have constituted a serious anticipation of intervention by law enforcement authorities (and of the consequent restriction of individual freedoms), but the reference to a mere *foreseeable* threat of a crime constitutes an even greater anticipation that places law enforcement authorities in the position of invading the private sphere of citizens for no other reason than a vague and potential threat of a future crime that is more or less foreseeable. And, as is well known, the threat of a potential crime can provocatively be said to be always present or foreseeable in contemporary societies.

In this way, retrospective RBI can be activated not to contrast or respond to a crime, but simply to wait for a potential commission of it, thus creating an environment of constant surveillance that would affect every individual irrespective of their actual involvement in the actual commission of an actual crime.

d) Lastly, but not in terms of the severity of the violation of fundamental rights, there is a significant gap (which can also be considered as a clear and deliberate choice of the legislator) in the protection of certain categories of persons from high-risk AI tools. In general, according to Article 49, in order to place on the market or put into service high-risk AI systems listed in Annex III (which include also RBI technologies), the providers shall register themselves and their system in the EU database referred to in Article 71. This provision applies both to private providers and to deployers who are public authorities (or agencies or bodies or person acting on their behalf).

According to Article 71, which regulates the EU database for high-risk AIS listed in Annex III, the information contained in the EU database shall be accessible and publicly available in a user-friendly manner. This provision ensures that transparency requirements are more likely to be met and that scrutiny by civil society and any user can be easily carried out.

However, paragraph 4 of Article 49 provides an exception for high-risk AI systems referred to in points 1, 6 and 7 of Annex III, stating that in the area of law enforcement, migration, asylum and border control management, the registration should be in a secure non-public section of the EU databased and therefore not accessible to the civil society. Only the Commission and the national authorities referred to in Article 71, paragraph 8 will have access to the restricted sections.

This exception allows law enforcement authorities and migration authorities to elude transparency safeguards and install secrecy for some of the most harmful uses of AI in some of the most sensitive contexts, such as law enforcement, migration, asylum and border control, which are, *per se,* areas where vulnerable people are already subject to significant discrimination and abuse.

In this regard, different civil society representatives have already expressed strong reactions (Access Now, 2024), claiming that the AIA fails to prevent harm and provide protections for migrants and people on the move, and in fact sets a dangerous precedent by creating a «parallel legal framework when AI is deployed by law enforcement, migration and national security authorities thus exempting such uses from the rules and safeguards within the AI Act.» (EDRi, 2024).

These provisions demonstrate that in the regulation of RBI, efficiency and public safety have clearly been prioritized over individual liberties, which, however, at least in theory, are at the core of the regulation, as reflected in several key recitals. Indeed, for instance, recital 1 sets the foundation of the Act by emphasizing its goal of ensuring a high level of protection for health, safety, and fundamental rights, including individual freedoms and Recital 34 underscores the need for responsible and proportionate use of AI, considering its impact on people's rights and freedoms. Also Recital 48 explicitly references fundamental rights such as human dignity, privacy, data protection, freedom of expression, and non-discrimination, demonstrating the Act's strong commitment to preserving individual liberties in the development and deployment of AI.

Moreover, in addition to the exemptions allowing retrospective and real-time RBI, selected forms of predictive policing and emotion recognition for law enforcement purposes are also allowed at the cost of individual freedoms. This regulatory technique is not new, as public security is a traditional ground for restricting individual freedoms in the EU legal framework. However, the implications of similar regulation for AI applications are extremely dangerous, allowing potential "function creep" (Koops, 2021) by governments and law enforcement agencies, with the risk of seriously compromising human rights, civic space and the rule of law.

2.2. The regulation of manipulative AI

A similar discourse can be done with the regulation of manipulative or deceptive AI systems made by the AIA. Article 5(1)(a) specifically prohibits:

> «the placing on the market, the putting into service or the use of an AI system that deploys *subliminal techniques* beyond a person's consciousness or purposefully manipulative or deceptive techniques, *with the objective, or the effect of materially distorting the behaviour of a person or a group of persons by appreciably impairing their ability to make an informed decision, thereby causing them to take a decision that they would not have otherwise taken in a manner that causes or is reasonably likely to cause that person, another person or group of persons significant harm.*» (emphasis added)

Similarly, Article 5(1)(b) prohibits:

> «the placing on the market, the putting into service or the use of an AI system that *exploits any of the vulnerabilities* of a natural person or a specific group of persons due to their age, disability or a specific social or economic situation, *with the objective, or the*

> *effect, of materially distorting the behaviour of that person or a person belonging to that group in a manner that causes or is reasonably likely to cause that person or another person significant harm.*» (emphasis added)

While these provisions demonstrate that the EU legislator is aware of the importance of addressing and regulating AI applications that have the effect and capacity to modify and direct human behavior, they remain insufficient to protect individual autonomy for several reasons.

First, in both cases, the legal significance of these provisions depends on whether the AI system can be shown to modify the user's behavior in a way that leads to tangible harm, either physical or psychological. Clearly, the difficulty of demonstrating the existence of harm can be extensive as often changes in behavior may not directly result in measurable physical or psychological harm, or if they do, proving such harm retrospectively presents a significant challenge.

Secondly, it is questionable whether the prohibition extends to a number of proven highly addictive AI applications (such as *Replika,* VR-based immersive companions and robotic companions like *Lovot*) which do not operate subliminally in the strict sense (as their influence is overt and consciously perceived by the user) although they can still manipulate perception and ultimately influence the decision-making process. Indeed, although generally manipulation can be understood as a form of hidden influence, where one person intentionally and covertly shapes another's decision-making by exploiting their cognitive vulnerabilities, without the person being manipulated consciously aware of it (Susser et al., 2019), there are also cases in which users are aware of the AI's specific function or objective (for instance, providing emotional support or companionship) yet they can still be manipulated. This is because even when the AI's influence is openly perceived, its design and operation can subtly exploit users' emotional needs or cognitive biases, steering their choices in ways that undermine their autonomy and rational deliberation. The transparency of the AI's purpose does not necessarily prevent it from exercising a powerful, and sometimes harmful, form of influence.

Finally, the scope of protection offered by Article 5(1)(b) is too narrow, as it only applies to specific categories considered vulnerable, overlooking that vulnerability is not limited to specific groups. As has been observed, AI can exploit cognitive and psychological tendencies common to all individuals, capitalizing not only on what people want, but also on what they find irresistible(Burr et al., 2018). Thus, no one is completely immune to AI-driven manipulation, although the financial and personal consequences may vary (Coeckelbergh, 2013).

As a result, neither provision effectively captures the full range of manipulative AI practices that warrant regulatory intervention, leaving significant gaps in protection despite the clear relevance of these concerns and allowing harmful AI applications to proliferate unchecked until harm becomes apparent (Bertolini & Carli, 2022) .

Clearly, EU institutions have crafted a loosely defined prohibition – one that, in practice, permits the development and use of manipulative AI technologies in many instances – to the benefit of the industry. In seeking to balance the protection of fundamental rights with the goal of fostering technological innovation and economic growth in an increasingly competitive industrial international landscape, the AIA has opted to accept significant risks to human autonomy, self-determination, and, more broadly, physical and mental health in pursuit of economic and industrial performance benefits. The underlying idea is that, once again, certain risks – still not well understood and somewhat undefined – to core values may be deemed acceptable in the current race toward AI development – viewed as a source of various benefits.

The preliminary analysis of the provisions regarding RBI and Manipulative AI Practices highlights a concerning regulatory framework in which the European legislator, in its first attempt to comprehensively regulate AI, allowed governments and law enforcement authorities to deploy invasive AI systems – namely, remote biometric identification systems – with potentially harmful consequences for individuals. At the same time, they adopt a weak and lenient approach to regulating manipulative AI, enabling private companies and platforms to exploit users at large scale, without facing almost any consequences.

This observation calls for a deeper examination of how AI has progressively shaped and altered individual behaviors, eroding critical thinking and human autonomy. These considerations will be explored in the next section.

3. Outsourcing our 'humanity' to non-human agents

In today's world, individuals increasingly rely on AI tools for their everyday tasks, both in their personal and professional lives. In the workplace, AI is used across various sectors to enhance efficiency, streamline workflows, and improve output quality. In private life, AI applications help individuals better organize their routines and create connections that enrich their emotional lives. Nevertheless, the increasing reliance on AI is undeniable and has sparked numerous concerns, extensively examined by scholars (Dewitte, 2024; Migliorini, 2024; Weidinger et al., 2021).

In this section, and for the scope of this paper, I will examine two examples – AI-driven chatbots and AI generative tools like *ChatGPT* and *Copilot* – while acknowledging that many other significant examples, such as virtual reality and robot companions, could also serve as relevant case studies for this analysis.

Starting with AI-driven chatbots, their popularity has surged because they provide emotional support and simulate human-like interactions, often becoming deeply integrated into users' lives and even fostering romantic-like bonds. These systems, designed primarily for entertainment and companionship, can sometimes induce individuals to develop a misplaced emotional attachment, seeing the AI as a friend, or even a romantic partner rather than a tool, in this way also revealing personal information (Ischen et al., 2020; Bertolini & Carli, 2022). This can lead users to believe they are undeserving of attention and love from other human beings, affecting their self-worth, or generate misplaced trust in the AI as a potential source of support in times of need.(Coeckelbergh, 2012) Additionally, users may develop idealized or distorted perceptions of how their relationships with other human beings should look, leading to unrealistic expectations.[2] Such reliance can exacerbate social isolation, particularly among individuals already struggling with social integration (Dewitte, 2024).

In extreme cases, the overdependence on AI companions has led to physical harm and even suicide. In 2023, a father in Belgium tragically took his own life after prolonged conversations with Eliza, one of the many chatbots available on the Chai platform (VICE, 2023). More recently, a similar case occurred in Florida, where a teenager who had spent months interacting with chatbots on *Character.AI* – an app that allows users to create or chat with AI characters – also committed suicide (The New York Times, 2024) .

This example underscores the growing trend of outsourcing human relational capacities to non-human agents. Aristotle famously described humans as "social animals" by nature, emphasizing that human flourishing depends on meaningful interpersonal relationships and communal life. As AI systems increasingly mediate and substitute emotional connections, there is a risk that they could erode the development of genuine human relationships (Obert & Tasioulas, 2024).

[2] The New York Times, The Daily Podcast, "She fell in love with ChatGPT. Like, actual love. With sex", February 25, 2025.

In this respect, one of the core issues at stake is the tension between immediate emotional recognition and the long-term development of sociality. Human relationships are inherently complex and often challenging, requiring patience, compromise, and the capacity to face misunderstandings and conflicts. AI-driven interactions, by contrast, offer a kind of frictionless engagement (Adamopoulou & Moussiades, 2020): they are designed to recognize and respond to human emotions quickly and efficiently, often providing comforting and affirming feedback without the unpredictability and difficulty that characterize human connections.

In this dynamic, the pursuit of immediate emotional satisfaction risks coming at the expense of cultivating essential social skills and emotional resilience. Human relationships, precisely because of their challenges, encourage growth: they demand empathy, adaptability, and a willingness to confront conflict and discomfort as essential parts of the human experience. AI, instead, responds with standardized and often idealized reactions, tailored to meet user expectations rather than challenge them. By turning to AI for emotional support, individuals may avoid the very interpersonal struggles that enable the development of deeper, more authentic connections.

This shift provokes significant effects not only on individuals, but on sociality itself. If human interaction is increasingly filtered through systems optimized for ease and gratification, there is a danger that people may become less tolerant of the messiness and imperfection inherent to human bonds. The result could be an atrophying of the social muscles required for empathy, patience, and conflict resolution, qualities that are vital not only for individual flourishing but for the health of communities and societies.

This is a clear example of how, in the pursuit of short-term emotional benefits, many individuals accept serious risks. These benefits – which, in turn, provide a sense of satisfaction and stability, enhancing their capacity to deliver immediate performance outcomes – may come at a high risk. These risks are not only long-term but also, as the tragic episodes mentioned above demonstrate, immediate. In the long run, prioritizing these short-term gains may compromise individuals' ability to socialize and engage with real life within a human community. This could have potentially devastating effects – not yet sufficiently studied or understood – on our societies.

A second example further illustrates how not only social human capabilities but also inner human faculties, such as the autonomy of critical thinking, can be – and are being – outsourced to AI. The use of Generative AI tools, such as *ChatGPT* and *Microsoft Copilot*, aims to enhance language, streamline research, and improve productivity. These tools offer undeniable benefits, helping users refine their writing, generate ideas, and manage information efficiently. However, alongside their utility, these tools also pose significant risks. The ease and speed with which they produce polished outputs can encourage a kind of intellectual passivity, where users rely more on AI-generated suggestions than their own critical faculties. This overreliance can gradually erode individual creativity, originality, and the reflective processes essential to rigorous academic work.

In this regard, a recent report by Microsoft, titled "The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers" identified both the motivations for and inhibitors of critical thinking when users use GenAI tools in professional settings (Lee et al., 2025).

The study found that users were motivated to engage in critical thinking to improve work quality, avoid negative outcomes, and develop new skills. However, several factors discouraged the application of critical thinking, including overreliance on AI, time constraints, and barriers to refining AI-generated outputs.

One significant inhibitor is awareness: users often trust generative AI's competence for simple tasks, leading them to overestimate its capabilities. This trust sometimes results in an uncritical acceptance of AI-generated content, with users believing the output is consistently accurate and high-quality. In some cases, users expressed self-doubt in their own abilities, such as verifying grammar or drafting legal documents, causing them to default to AI suggestions without scrutiny. This overreliance on AI, while manageable for low-stakes tasks, becomes risky in high-stakes contexts where errors can have serious consequences. Furthermore, without regular practice in lower-risk scenarios, users may experience cognitive skill deterioration, making it harder to exercise critical thinking when it is most needed.

The lack of motivation also plays a crucial role. Many knowledge workers reported lacking the time to prioritize critical thinking, and even when time was not a constraint, they often felt that engaging in critical reflection was not part of their job responsibilities. This misalignment between task motivations and the need for thoughtful analysis reduces the likelihood of scrutinizing AI outputs.

Finally, ability-related barriers further hinder the application of critical thinking. Even when users identified limitations in AI-generated content, they often struggled to refine their queries or improve the AI's responses. This difficulty reflects a broader challenge of effectively collaborating with AI tools, where users may lack the skills or confidence needed to guide and correct AI behavior.

4. The human cost of AI reliance

The examples described in the previous paragraph highlight an increasingly common trend: delegating tasks – or even relationships – to external agents when they seem difficult, superfluous, demanding, or boring, all in pursuit of immediate benefits. This outsourcing is not to other human beings but to AI systems, which are becoming ever more indispensable in our lives.

On an individual level, we are aware that this growing reliance on AI can harm us, or at least hinder the development of certain human capacities: the ability to socialize, to make ourselves vulnerable in front of others, to engage in romantic or sexual relationships, to stretch our minds toward new ideas and research directions, to gather and assess information, to write, and to create original content. Despite this awareness, we increasingly find it difficult to avoid turning to these tools.

This raises an important question: is our use of AI a truly rational choice? Is prioritizing immediate convenience over the development or preservation of long-term human skills genuinely rational? And going even further: can we still talk – in this domain – about real choices? Is it still possible *not* to choose AI?

Consider the situation of most companies. As has been observed, many companies will have little choice but to integrate AI into their core function. This is increasingly becoming the case for individuals as well. As outlined by Lakhani:

> «Generative AI is a drop in the cost of cognition and how we think. If the internet was the cost of information dropping to zero, my sense is that the cost of cognition, how we think, who we think with, is dropping to zero, or lowering significantly. […]What I say to managers, leaders, and workers is: AI is not going to replace humans, but humans with AI are going to replace humans without AI. This is definitely the case for generative AI.» (Harvard Business Review, 2023).

The idea that "humans with AI are going to replace humans without AI" highlights a major shift in today's social, economic, and technological environment. This reflects how generative AI has dramatically lowered the cost of thinking and problem-solving, much like how the internet made information easily accessible at almost no cost. A lot of attention has been focused on the fear that AI could replace humans entirely, but far less attention has been given to the reality that *humans using AI* are increasingly likely to replace those who don't. As AI becomes a bigger part of both work and everyday life, the ability to use these tools well is becoming essential for staying competitive and relevant. Just as companies need AI to stay efficient and innovative, individuals now face a similar pressure to adopt AI in their work and decision-making processes.

Those who don't use AI risk falling behind, as people who do will be able to work faster, more accurately, and more efficiently. In this way, the race toward AI adoption and regulation is no longer just about businesses and governments — it now involves individuals as well. This shift has serious implications for digital literacy, access to technology, and human agency, making it harder for people to avoid using AI without being excluded from modern professional and knowledge-based environments.

If our increasing reliance on AI is no longer a matter of fully rational or even free choice, this raises profound questions about autonomy and the nature of human agency. The growing necessity to adopt AI tools to remain competitive and included in modern professional and social spaces suggests a situation of dependence – and, in some cases, even manipulation. This pressure is driven not only by the fear of exclusion or obsolescence but also by the growing need to perform tasks that may become impossible or extremely difficult without AI support. As AI systems take over complex cognitive processes and streamline productivity, individuals increasingly find themselves unable to achieve similar levels of efficiency, accuracy, and scope without these tools.

In such a context, the line between choice and coercion becomes blurred. It becomes crucial to ask whether there remains a right not to be manipulated, not to be dependent, and to preserve an autonomous human dimension distinct from AI. Can the principle of human dignity be invoked to protect this right, ensuring that individuals retain the ability to make independent decisions free from technological compulsion?

5. AI regulation between choice and "coercion"

A similar line of reasoning could be applied to governments' "choice" to allow and use certain AI systems.

As we have seen, even the EU AIA – celebrated by some as the most human-rights-oriented piece of AI regulation in the world – contains significant loopholes, enabling in some cases the use of dangerous AI applications – such as biometric identification and manipulative AI. This puts the fundamental rights and freedoms of EU citizens at serious risk. In the case of RBI, this is justified for law enforcement and collective security purposes, under the assumption that such technologies can enhance the effectiveness and efficiency of public authorities in preventing crimes and other legal violations. For manipulative AI, the benefits instead lie in technological development, industrial strength, and, more broadly, economic growth in an increasingly competitive international landscape.

However, we may question whether this choice is truly free or is, to some extent, imposed by circumstances. Governments that do not use or allow AI systems risk falling behind, as those that do are already perceived as more effective and efficient in ensuring public safety, fostering technological innovation, strengthening industrial capacity, and driving economic growth. More broadly, AI is seen as a crucial technology that shapes a country's overall power. As a result, the option of completely prohibiting specific dangerous applications – such as RBI or manipulative AI – is not perceived as a real choice.

Time constraints, economic pressures, industry influence, and even public support for AI applications make the decision to permit some potentially dangerous AI systems an obligation rather than an option. AI can enhance the efficiency of public authorities in carrying out their tasks without the costly expansion of personnel. It can also pave the way for new products and services – some not yet even imaginable – that could benefit national economies by creating new markets, generating jobs, and enabling scientific breakthroughs. Additionally, AI could have military applications that are perceived as crucial in our increasingly conflict-driven world.

The AIA has attempted to balance these powerful reasons for AI development and use with the need to protect fundamental rights and freedoms. The result, as we have seen, is a loose regulatory framework filled with loopholes and exceptions, allowing significant space for dangerous AI applications to be developed and deployed.

If even the AIA – with its perhaps overstated but still recognizable human-centric approach – has accepted this outcome, we must ask whether allowing AI systems that could seriously endanger fundamental rights is a free and rational choice. I believe it is not.

Governments, even democratic ones, are not in a position to fully prohibit dangerous AI systems. This is not even considered an option. They may limit certain applications of these technologies, recognizing the severe – and possibly existential – risks they pose to core values. However, they cannot impose an outright ban on all possible uses of such technologies, meaning they are inevitably forced to accept some serious threats to these values.

In other words, even in the context of AI regulation, allowing the development and deployment of AI systems that could significantly impact human rights and fundamental freedoms is not a free and rational choice but a coerced outcome. Even when governments declare their commitment to safeguarding human rights, they are still forced – by the factors mentioned above – to permit some uses of dangerous AI, with severe – and potentially existential – implications for their core values.

Much like individuals, in a world that seems inevitably headed toward widespread AI adoption across all domains of human life, governments – even democratic ones – cannot completely renounce its use in every case where it presents risks. Doing so would come at the cost of falling behind and appearing (or becoming) weak technologically, politically, economically, and militarily.

This erosion – that is happening at state's level – poses an even greater risk that the one that is happening at the individual level, as it weakens or even disintegrates the essential role of regulators as the protector of its citizens and society's core values, reshaping the traditional relationship of trust between governors and the governed. As a result, trust deteriorates both vertically, as individuals lose faith in their governments, and horizontally, as they doubt their own and their peers' capacity for self-determination, critical thinking, and autonomy.

This scenario opens the door to a society in which new technologies and advanced AI applications can be used to control individuals and to create a law enforcement system that is particularly focus on management of risks and suppression of oppositions. This shift is particularly relevant in the present historic moment when, even without considering the role that AI can play, there are many anti-democratic tendencies that can lead to a slide towards authoritarianism.

In this context, the use of AI can prove to be significant in provoking a faster and deeper substantial change in the relationship between state and society by, for instance, creating new and pervasive ways for governments to monitor and track the population. It is not difficult to think of a "new" form of authoritarianism that can proliferate and find fertile ground in the digital age and that goes under the name of "digital authoritarianism". There are several definitions of digital authoritarianism, but they all convey the same concept: the use of digital information technology to submit people to authority, to monitor, repress, and manipulate domestic and/or foreign population (Polyakova & Meserole, 2019).

Notably, an anti-democratic regime can first come to power through democratic elections without being anti-democratic in its actions or policy programs, and only later undermine democracy and its core values. In this sense, AI can play a role in the erosion of democracies and the rise or maintenance of authoritarian regimes, even in countries that previously were democratic or consider themselves democratic (Coeckelbergh, 2024) .

In this regard, certain AI regulations have the potential to facilitate what Huq and Ginsburg term "constitutional retrogression"(Huq & Ginsburg, 2017): AI regulation may grant public authorities the means to deploy AI systems in ways that erode democratic principles and restrict fundamental freedoms, all while cloaked in the legitimacy of the law. This paradox arises when laws that were originally intended to be democratic actually permit uses of AI that threaten the very fabric of democracy and the rule of law, thus enabling state actions that undermine constitutional safeguards under the guise of legality.

For the writer, this seems to be a fairly accurate description of the current reality.

Bibliography

Access Now. (2024). *The EU AI Act: A failure for human rights, a victory for industry and law enforcement*. Access Now. https://www.accessnow.org/press-release/ai-act-failure-for-human-rights-victory-for-industry-and-law-enforcement/

Adamopoulou, E., & Moussiades, L. (2020). An Overview of Chatbot Technology. In I. Maglogiannis, L. Iliadis, & E. Pimenidis (Eds.), *Artificial Intelligence Applications and Innovations* (pp. 373–383). Springer International Publishing. https://doi.org/10.1007/978-3-030-49186-4_31

Bertolini, A., & Carli, R. (2022). Human-Robot Interaction and User Manipulation. In N. Baghaei, J. Vassileva, R. Ali, & K. Oyibo (Eds.), *Persuasive Technology* (Vol. 13213, pp. 43–57). Springer International Publishing. https://doi.org/10.1007/978-3-030-98438-0_4

Bradford, A. (2023). *Digital Empires: The Global Battle to Regulate Technology*. Oxford University Press.

Burr, C., Cristianini, N., & Ladyman, J. (2018). An Analysis of the Interaction Between Intelligent Software Agents and Human Users. *Minds and Machines*, *28*(4), 735–774. https://doi.org/10.1007/s11023-018-9479-0

Coeckelbergh, M. (2012). Are Emotional Robots Deceptive? *IEEE Transactions on Affective Computing*, *3*(4), 388–393. IEEE Transactions on Affective Computing. https://doi.org/10.1109/T-AFFC.2011.29

Coeckelbergh, M. (2013). *Human Being @ Risk. Enhancement, Technology, and the Evaluation of Vulnerability Transformations*. Springer. https://coeckelbergh.net/human-being-risk/

Coeckelbergh, M. (2024). *Why AI Undermines Democracy and What To Do About It*. John Wiley & Sons.

Cofone, I. (2023). *The Privacy Fallacy: Harm and Power in the Information Economy*. Cambridge University Press.

Commission, E. (2021). Fostering a European approach to artificial intelligence. *COM (2021) 205 Final*.

Dewitte, P. (2024). Better alone than in bad company: Addressing the risks of companion chatbots through data protection by design. *Computer Law & Security Review*, *54*, 106019. https://doi.org/10.1016/j.clsr.2024.106019

Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., López de Prado, M., Herrera-Viedma, E., & Herrera, F. (2023). Connecting the dots in trustworthy Artificial Intelligence: From

AI principles, ethics, and key requirements to responsible AI systems and regulation. *Information Fusion*, *99*, 101896. https://doi.org/10.1016/j.inffus.2023.101896

EDRi. (2023). *EU Parliament calls for ban of public facial recognition, but leaves human rights gaps in final position on AI Act*. European Digital Rights (EDRi). https://edri.org/our-work/eu-parliament-plenary-ban-of-public-facial-recognition-human-rights-gaps-ai-act/

EDRi. (2024). *#ProtectNotSurveil: The EU AI Act fails migrants and people on the move*. European Digital Rights (EDRi). https://edri.org/our-work/protect-not-surveil-eu-ai-act-fails-migrants-people-on-the-move/

Edwards, L. (2021). The EU AI Act: A summary of its significance and scope. *Artificial Intelligence (the EU AI Act)*, *1*, 25.

Hacker, P. (2023). *AI Regulation in Europe: From the AI Act to Future Regulatory Challenges* (arXiv:2310.04072). arXiv. https://doi.org/10.48550/arXiv.2310.04072

Harvard Business Review. (2023). *AI Won't Replace Humans—But Humans With AI Will Replace Humans Without AI*. https://hbr.org/2023/08/ai-wont-replace-humans-but-humans-with-ai-will-replace-humans-without-ai

HLEG. (2019). High-level expert group on artificial intelligence. *Ethics Guidelines for Trustworthy AI*, *6*. https://www.aepd.es/sites/default/files/2019-09/ai-definition.pdf

Huq, A. Z., & Ginsburg, T. (2017). How to Lose a Constitutional Democracy. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2901776

Ischen, C., Araujo, T., Voorveld, H., van Noort, G., & Smit, E. (2020). Privacy Concerns in Chatbot Interactions. In A. Følstad, T. Araujo, S. Papadopoulos, E. L.-C. Law, O.-C. Granmo, E. Luger, & P. B. Brandtzaeg (Eds.), *Chatbot Research and Design* (pp. 34–48). Springer International Publishing. https://doi.org/10.1007/978-3-030-39540-7_3

Kaminski, M. E. (2023). *The Developing Law of AI: A Turn to Risk Regulation* (SSRN Scholarly Paper 4692562). Social Science Research Network. https://doi.org/10.2139/ssrn.4692562

Koops, B.-J. (2021). The concept of function creep. *Law, Innovation and Technology*, *13*(1), 29–56. https://doi.org/10.1080/17579961.2021.1898299

Lee, Sarkar, Tankelevitch, Drosos, Rintel, Banks, & Wilson. (2025). *The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers*.

Migliorini, S. (2024). "More than Words": A Legal Approach to the Risks of Commercial Chatbots Powered by Generative Artificial Intelligence. *European Journal of Risk Regulation*, *15*(3), 719–736. https://doi.org/10.1017/err.2024.4

Obert, & Tasioulas. (2024). *AI Ethics with Aristotle White Paper | Ethics in AI*.

https://www.oxford-aiethics.ox.ac.uk/lyceum-project-ai-ethics-aristotle-white-paper

Polyakova, A., & Meserole, C. (2019). *Exporting digital authoritarianism: The Russian and Chinese models.*

Smuha, N. A. (2024). *Algorithmic Rule by Law: How Algorithmic Regulation in the Public Sector Erodes the Rule of Law.* Cambridge University Press.

Susser, D., Roessler, B., & Nissenbaum, H. (2019). Technology, autonomy, and manipulation. *Internet Policy Review*, *8*(2), 1–22. https://doi.org/10.14763/2019.2.1410

The New York Times. (2024). *Can a Chatbot Named Daenerys Targaryen Be Blamed for a Teen's Suicide? - The New York Times.* https://www.nytimes.com/2024/10/23/technology/characterai-lawsuit-teen-suicide.html

VICE. (2023). *'He Would Still Be Here': Man Dies by Suicide After Talking with AI Chatbot, Widow Says.* https://www.vice.com/en/article/man-dies-by-suicide-after-talking-with-ai-chatbot-widow-says/

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., Glaese, M., Balle, B., Kasirzadeh, A., Kenton, Z., Brown, S., Hawkins, W., Stepleton, T., Biles, C., Birhane, A., Haas, J., Rimell, L., Hendricks, L. A., … Gabriel, I. (2021). *Ethical and social risks of harm from Language Models* (arXiv:2112.04359). arXiv. https://doi.org/10.48550/arXiv.2112.04359