

DOI: 10.24193/tras.SI2025.3  
Published First Online: 12/16/2025

## CRITICAL REFLECTIONS ON A EUROPEAN APPROACH ON HUMANS AND MACHINES

Emil BOC

**Emil BOC**

Associate Professor, Political Science Department, Faculty  
of Political, Administrative and Communication Sciences,  
Babeş-Bolyai University, Cluj-Napoca, Romania  
E-mail: boc@fspac.ro

### **Abstract**

Artificial Intelligence (AI) is not a neutral tool. The debate on AI is not about technology, but about power, order and values. It is desirable to take a multidisciplinary approach and place technology within democratic ethical and political norms. The European Union has built a coherent and ambitious regulatory framework, backed by a clear political vision, a liberal one, based on rules and procedures, on the one hand, and respect for individual rights and freedoms, on the other.

**Keywords:** Artificial Intelligence (AI), power and values, Ethical Artificial Intelligence, democracy, European Union.



*Transylvanian Review  
of Administrative Sciences,*  
Special Issue 2025, pp. 49–68

## 1. Introductory remarks

Artificial Intelligence (AI) is probably today's trendiest topic for experts in various fields. In the spring of 2025, news appeared about how OpenAI's GPT-4.5 passed the Turing test. The principle of the test is simple: a human asks questions to another human and to a machine and has to decide who is a human and who is a machine. The study conducted by Cameron Jones and Benjamin Bergen at the University of California San Diego showed that in 73% of the cases the AI was indistinguishable from the human (there were 284 participants). The researchers who conducted the study point out that they do not support the idea that the test is enough to talk about human intelligence, but rather an 'imitation of human intelligence' (Assaad, 2025). This study has given a new impetus to the multi-layered and interrelated debates about AI and AI's relationship with humanity.

Technology seems to be coming upon us, totally changing the game, the opportunities and especially the threats are here. It seems obvious that AI is not a neutral instrument as many would like to see it. It is time to remember what Heidegger (1977, p. 12) said: 'technology is a way of revealing' (p. 12). And what the Greeks meant by 'technê' (τέχνη or τέκτων, Parry, 2024), a word with multiple meanings, lost perhaps for a while, and in need of being rediscovered today. 'Technê' means knowledge, in general, it also means episteme, concrete realization, empirical approach, skill, and it also means art – 'because the essence of technology is nothing technological, essential reflection upon technology and decisive confrontation with it must happen in a realm that is, on the one hand, akin to the essence of technology and, on the other, fundamentally different from it. Such a realm is art.' (Heidegger, 1977, p. 35). Trying to understand technology means looking far beyond the technical aspects.

The new revolution coming upon humanity is forcing us to make decisions and take actions, which may be more or less wise. In this context, the European Union has formulated responses based on its own values (through all its institutions). In this paper, I set out to briefly analyze how the position of the European Committee of the Regions (CoR) (2024) fits into the global debate in general, and the European Union's approach to AI in particular. It should be noted that the CoR formulates recommendations on the basic principles of the European approach, emphasizing the role of local and regional authorities in the governance system of AI, because these authorities are closest to the people, they solve most of the problems that arise in real life.

## 2. About humans and machines

### *2.1. A discussion about what we want the world we will live in to look like*

Artificial intelligence (or AI) has long been just an academic field, but AI is a term used in a broad sense describing a set of products, techniques and services related to computer science with major implications in our lives, with threats and opportunities (Harding, 2024). In other words, AI is a result of technological development, simply put 'a machine',

to use an easy-to-understand word that has entered pop culture. Instead of the term AI, the term ‘machine learning’ is sometimes used, which is a specific method by which algorithms are fed (trained) with data to build predictive models. Starting from the distinction proposed by John Searle (1980) between ‘weak AI’ (AI as a tool that becomes more and more complex but remains a tool) and ‘strong AI’ (what would be called AGI, or artificial general intelligence), *i.e.* a machine that could think, that can do everything a human does and even more. Vallor (2024, pp. 21–26) points out that ‘narrow AI’ is most appropriate for the current state of AI, given that the tools available today are not ‘underpowered’ and the way machine learning systems perform specific functions is very good. But they remain below the level of AGI (it is not clear if and when it will be reached, nor whether it is a good idea to reach it, despite the collective imagination and AI marketing), because they only provide statistically plausible answers.

Bill Gates (2023) uses the term ‘agents’ for the collections of algorithms that are just changing our future. The term is interesting because it reminds us of a movie character, Agent Smith in *Matrix*, the ‘agent’ who tends to occupy all possible worlds.

There is an evidence-based consensus that artificial intelligence is already having a major impact and will have a major impact in education, science, finance, social networking and manipulation, manufacturing (where it has already been playing a major role for many years), labor market structure, surveillance, military, etc.

This development raises many questions, and the answers require a multidisciplinary approach that goes beyond the technical framework and builds on it. In addition to the multidisciplinary approach, two closely interlinked aspects emerge as extremely important: mankind’s relationship with AI (and the political dimension that stems from this) and fear (fear of machines, fear of having lost our compass, fear of not knowing who we are). In short, AI has ceased to be a topic for expert-level discussion and has become a topic that forces us, once again, to think about what we want the society we will live in to look like. These developments are accompanied by a lot of fear and a fascination with techno-utopias and techno-dystopias in which artificial intelligence is not a tool but a creature. For example, in movies, the most influential artistic medium of the last century, or in science fiction literature, fear and fascination are evident; we may have fallen in love with our machines (Danesi, 2024, pp. 56–57). I mention just a few examples: *2001: A Space Odyssey* (Stanley Kubrick), *Blade Runner* (Ridley Scott), *A.I.* (Steven Spielberg), or Asimov’s stories. On the other hand, if one looks more closely, these works are not about AI; they are about the human condition.

Within this framework, moving towards what is meant by a European approach, the Committee of the Regions (2024) calls for some seemingly trivial but of utmost importance things: human control over AI, respect for humanity and an ethic of AI use, sustainability of AI use, listening to the voice of stakeholders, respect for privacy, ‘individual and collective rights’ (p. 5), ‘workers’ rights’, ‘fair and equitable outcomes for all individuals and communities’ (p. 6), also calling for attention to ‘business models and market fit’

(p. 5). This is a clear political vision, a liberal one, based on rules and procedures, on the one hand, and respect for individual rights and freedoms, on the other. We should not pretend that the discussion on AI is a technical one. It is basically a political discussion about what we want the world we will live in to look like, about order in society, and about rights and freedoms. AI is not just a technical question, but also one about power and politics (Coeckelbergh, 2022). To create an ‘ethical design’ for AI, we need a holistic, multidisciplinary approach (Agerwala, 2024, pp. 66–68), as the development of AI is not driven by ethical principles, but by market forces and power interests.

Why ‘fair and equitable outcomes for all individuals and communities’? Let us go back in time and remember that the first machine that had an impact on its environment and could make adjustments was the one developed by James Watt at the time of the industrial revolution (Joshi, 2024, p. 113) and it was based on steam (at its core, AI is nothing new, it’s just that we’re talking about it now because the tools have become extremely powerful). Today, we are facing a revolution that can be linked to the industrial revolution, and as with previous revolutions, if solutions to people’s problems are not found, the shocks caused by new technologies can destabilize the social order and bring forward radical alternative solutions (Garfinkel, 2024, p. 86). The totalitarianisms of the 20<sup>th</sup> century bear witness. A 21<sup>st</sup>-century totalitarianism would be a much more elaborate communism or fascism, which would make the old odious regimes seem medieval, brutal and effective in their time to annihilate humanity, but totally outdated, because they would have ubiquitous digital surveillance and almost infinite capacities for social manipulation at their disposal.

As for the future of war, artificial intelligence has exceptional potential (Dafoe, 2024). The future of warfare is not about some guys sitting in a basement commanding drones like in a computer game (that’s the present!), the future is a robot maintaining a network of robots commanding other combat robots. Man’s relationship with technology is one that has its progress intertwined with organized violence. The more ‘intelligent’ the machines are created, the greater the destruction. The more combat automatons are equipped with artificial intelligence (albeit in a weak sense), the more thorough and precise the destruction. And here I don’t mean the possibility of AI taking over, just humans. Simply put, machines create extreme versions of humans (Joshi, 2024, p. 111) in various respects, including (or especially) in that of organized violence. Today, we see organized violence trampling on individual rights and freedoms, and collective rights. The answer to the question of what we will do as AI becomes more sophisticated is most likely more powerful weapons. A global approach to control this kind of development is easy to demand when no regime challenges order by force, but we see that the most cynical realism has returned. But it is important, at least at the European level, to meet security needs without losing citizens’ rights and freedoms.

Equitable outcomes, workers’ rights, business models and market fit, socio-economic issues are also important. Let’s consider the following simple example: when supermarkets

replace employees, who used to work at the cash registers, with machines, it's not because people were weaker or slower, it's because they replaced some of the work with the effort of customers, who have to get into queues according to how many products they buy, then have to scan them themselves, if the barcode is not visible they have to enter it manually – exactly what the laid-off employees used to do. Beyond the fact that it is not clear what the fired employees are going to do – will they write the code? – here we see only a small part, like a one-second glimpse behind the scenes of a show, of the power differential between citizens and corporations. Corporations can force you to work for them and even be happy. Here is a small illustration of what Vallor (2024, pp. 199–200) calls 'the two most destructive trends of twenty-first-century digitalization'. The first is the transfer of power and wealth into fewer and fewer hands. The second is harder to perceive and reinforces the first: the erosion of moral and political trust between people and the erosion of trust in ourselves, because we are told that machines are better than us in almost every sense. It is possible that the processes that are in place today do not let certain disadvantaged groups to have their voices heard and thus AI produces 'epistemic injustice', because being part of the power structure does not let these voices to participate in shaping 'collective moral and social life' (de Sio, 2024, pp. 178–179). Machines rely on the fact that our digitally mediated behaviors are predictable. This is a major business opportunity for their owners, and the world thus constructed is poor not by nature but by design; and they are 'habitus machines' (Romele, 2024) because they produce classifications of people – usually based on previous human-made classifications embedded in algorithms.

Machines are not public goods; they are private goods that shape the world. Their economic success has a limit. The limit of the 'tech revolution' was also seen in the dot.com crash – a development based on the mirage of the internet and greed (greed is what you can count on when it comes to people). The offer seems generous – AI means innovation, competitiveness, etc., but sustainable innovation does not mean cutting social costs (Vallor, 2024, p. 56). How do we not lose the social component – in other words, how do we close the gap between the haves and have-nots, when it seems that the haves don't really want to give anything to the have-nots. What is needed is a social perspective and a set of public policies that counteract the labor market effects without introducing distortions in the economy (Korinek and Stiglitz, 2017).

The adverse effects of technical progress are inseparable from the positive effects, and man is at the center of these processes. In real life, in immediate empirical experience, man is forced to experience these effects (Ellul, 2008). It is therefore always worth asking ourselves, when making a decision about technology, whether or not the effects end up contradicting the values for which we advocate progress. Is the development of AI leading us towards serfdom, authoritarian or even totalitarian societies? – is a question raised by several thinkers. As the institutions of force incorporate AI into their tools and as their budgets grow, the voice of the citizens is increasingly muted, allowing the institutions of force to tell us what it means to be safe. We are allowing them, sometimes even asking them

– as we see too many campaigns to promote fear and create fault lines in society – to be more oppressive. Technology has the potential to maintain order, but which order it maintains is a matter for citizens to determine, because any power becomes abusive if given the opportunity. AI systems are used to exercise power (Lazar, 2024). For example, citizen surveillance has a profound impact on the way we behave, both in social life in general and in the digital space in particular because of fear (O’Shea, 2019, pp 50–51). Online discussions are changing, people are avoiding certain controversial topics and ideas, and freedoms are shrinking, and this carries over into other aspects of public communication and social life. Power disciplines us, as always, under the pretext of the common good. Those who lived through the totalitarianisms of the 20<sup>th</sup> century remember fear and doublespeak well.

AI can contribute to the realization of a ‘disciplinary device’ in Foucault’s (1975) sense, in which the individual is supervised and distributed in a certain category, and in which people come to submit to power without the need for repressive authority. We see coming closer to us two fundamental political dreams of power exercised against people’s possibilities of contact – that of the ‘leper’ and that of ‘stopping the plague’, that is, the dream of a ‘pure community’ and that of a ‘disciplined society’ (Foucault, 1975, pp. 199–201), two different and not incompatible but complementary dreams. We see today the dream of pure communities expressed in campaigns of disparagement and demonization of groups of people waged on social networks with the major involvement of AI. We then see how AI is being used to surveil citizens and to reward or punish later. There are multiple possibilities to create a ‘totalitarian dynamic’ (Coeckelbergh, 2022, p. 115).

The question arises whether AI can protect individual freedom and democracy. I remember that the explosion of the Internet was seen in mainstream discourses in the public sphere as an opportunity to change the way politics is done, to bring greater citizen participation, to change processes and power structures, to give voice to the oppressed and disadvantaged, and to give meaning and strength to fundamental human values. Today, things are viewed with concern rather than optimism.

Then I remember the enthusiasm for using social networks for the same things. For example, during the Arab Spring, where today we have largely dictatorship or chaos, they certainly helped people to organize, but later they helped the authorities identify them. Today, social networks are not a space for deliberation – in Habermas’ sense – but rather means of control and manipulation. Algorithms gather data about you and get to know and understand you better and better, and then give you what you want. Instead of a space for debate, bubbles or echo chambers (Coeckelbergh, 2022, p. 105) emerge, in which you are not exposed to different opinions or content you dislike, but to things that entertain you and reinforce your opinions.

All over the world, we see a shift towards authoritarian regimes, and this evolution cannot be separated from the use of AI. Even in democratic countries, we see developments that give rise to questions (the answers to which we do not like) regarding the use of social networks that are managed by their nature and allow themselves to be managed by AI.

And attempts at stereotypical answers that blame everything on the ‘enemy’, therefore answers that go in the direction of a ‘pure society’, are not correct answers. Of course, they are not correct answers if we still want individual freedom and democracy. The fear is therefore not unjustified, AI is a tool that can be used against freedom and against democracy.

As we have shown, the idea of AI neutrality is an illusion. Technology is always more than its material components, it is closely linked to symbolic forms, it is linked to power, and thus to politics. The issue of rules concerning AI is primarily a political-socio-economic one and tech just has to adapt (Weissingner, 2024), in other words, it is a question of value-based political decision-making. We need politics of artificial intelligence (Romele, 2024) and public policies that derive from the political vision. I’ll give you two examples. Let’s remember, first, that at the heart of innovation is the fact that something has to be created, produced, sold and used, and this product may include some, but exclude others (even brutally). The consensus reached in the public sphere about AI, the kind of consensus on which democracy is based, legitimizes some actors and their discourses and excludes others, simply put, the poor and marginalized (Romele, 2024, pp. 163–168). In the kind of discourse dominant at the moment in the public sphere, we hear more and more about the benefits of AI. We are told, directly or more often indirectly, that humans are slower, weaker, less rational, less capable, less valuable, and above all more expensive than AI. And then the elites tell us to build a world where we trust each other and ourselves, but do we really believe that? Second, an example with brutal effects in immediate reality: as hardware has gotten more and more powerful, it consumes more and more power. Have enough energy and at a reasonable price, you can be a competitor in AI. This element, the energy required, adds further pressure on political decisions and hence public policy, whether we recognize it or not.

## *2.2. Between parallel mirrors*

ChatGPT and similar systems are the stars of the present, with GPT (Generative Pre-training Transformer) being a machine that mimics a human-like conversation and generates text (grammatically correct, semantically coherent and, most of the time, plausible, but not necessarily ‘true’, in the sense that ‘true’ would mean an empirically verifiable correspondence between the states of fact and the sentences emitted by GPT). Perhaps it may seem like magic in a world where education matters less and less, and writing a coherent text seems a burden for more and more people. But it’s nothing new in principle, the first model that tried to do the same thing was MIT’s Eliza in the 1960s. Eliza was followed by Alice in the 1990s (Joshi, 2024, p. 143) and then OpenAI launched GPT in 2022. New, however, is the power of the algorithms that give this ‘toy’ the ability to generate interaction in a natural language as well as in a multimedia language, following training on large databases. What’s more, ChatGPT has been made available to users for free, and other tools of the same type have been brought to market by various companies. These are some superb examples of AI, but I don’t consider them to be what is called AGI

– artificial general intelligence (AGI) that surpasses human intelligence, even though the horizon seems close, and we are always threatened with it.

ChatGPT amazes us, but it only mimics speech, it mimics it better and better since Eliza, being a tool made up of algorithms that extract information from human artifacts and behaviors and provide predictions, and if it provides incorrect answers or prescribes incorrect behaviors it is because the algorithms have selected the expression of incorrect opinions and attitudes that we already have (Vallor, 2024, p. 43), the problem is the ‘food’ we provided.

We fear the superiority of machines. Before we get that far, there are already plenty of reasons to be afraid. We have already understood that ChatGPT can lie, in the sense that it says plausible but false things. Moreover, about the deepfakes that can be generated, actor Jordan Peele warned us back in 2018, demonstrating that AI can show President Obama delivering a message written by Peele (Buckner, 2024, pp. 58–59). Here we have a concrete example of a direct threat to the space of political deliberation and, from there, to democracy, as control over certain algorithms can bring to prominence discourse that is fascist or communist, far from any idea of democracy, and, on the other hand, the prominence of power discourse (whatever that may be). Alan Turing stated: ‘At some stage therefore we should have to expect the machines to take control’ (circa 1951, p. 475), however he said this in the context of agreeing with Lady Lovelace<sup>1</sup> that the machine ‘has no pretensions whatever to originate anything’, but ‘can do whatever we know how to order it to perform’ (1951, p. 482). This fear does not exclude the one related to the control of a superintelligence, but is complementary and is present here. There is no need for superintelligence to relinquish control to algorithms and thus produce dehumanization. Because relinquishing control to algorithms, specifically relinquishing control over the food they chew and deliver back to us, actually means relinquishing control to corporations and governments. Essentially, ownership of machines is and will remain a central issue; any discussion that does not take this aspect into account deviates from empirical validation. We see how today major players rally and change their policies according to the shifting power, and that does not help at all when it comes to rights and the common good.

Going further, the question arises whether or not machines have/will have all the characteristics that make us human. Depending on the empiricist or nativist approach, deep learning systems are ‘at least slightly conscious’ and rapidly moving towards AGI, or they can only imitate human intelligence without capturing the essence of humanity (Buckner,

---

1 I would like to mention in parentheses that Augusta Ada King, Countess of Lovelace, born Byron, was the first programmer in the world, and together with Charles Babbage, what we can call, ‘the first computer geeks’. She made a fundamental contribution by showing that ‘by manipulating symbols according to rules, any kind of information, not only numbers, can be operated on by automatic processes’, 100 years before the construction of the first computer (O’Shea, 2019, pp. 119–122).

2024). Most likely, a moderate approach, with moderation seen as a fundamental virtue, is the right one.

I mentioned the experiment in which GPT-4.5 passed the Turing test in 73% of cases. However, is the Turing test sufficient? A pragmatic and consistent argument in its favor is that if the majority of people attribute the ability to think to other people they converse with, it is reasonable to attribute thinking to an artifact that converses well enough to make us believe it is a human (Crockett, 1994, p. 200). Moreover, even if there are reservations about whether this test is relevant or not, the things that people believe to be true are true in their consequences (Thomas's theorem).

On the other hand, there are objections related to the Turing test, and Buckner (2024, pp. 75–77) shows that attempts to make a quantitative and qualitative assessment of AI in relation to human intelligence are controversial due to their experimental validity, despite the fact that such assessments are currently being made on humans.

Another perspective that deserves to be mentioned is the one that starts from the premise that the physical system called human body and intelligence are not distinct things, but human bodies are intelligent (Susser, 2013, pp. 285–286). 'Hardware' and 'software' cannot be understood as being clearly distinct when we talk about people. Thus, the very structure of physical systems should be designed to produce intelligent behavior. Moreover, the human body has functions that are not present in digital agents, which are forced to formalize everything, which is an 'infinite task' (Dreyfus, 1972, p. 167). No dataset can contain the world; the uncertainty principle has long shattered a deterministic vision of the world. This leads to the idea that there is no danger of machines becoming humans based on 'self-improvement' – although there are researchers who claim that if 'self-improvement' is sufficiently sophisticated, it could lead to an intelligence explosion (Hägström, 2022, pp. 156–159), but I have shown that there are plenty of other ways in which algorithms dehumanize us and can destroy democracy.

AI systems are built as 'mirrors of human intelligence' (Vallor, 2024), because they feed on our needs, our expectations, our decisions, our values and are thus inseparable from humanity, yet they remain just structures that seem human because they perform the same tasks as humans without having consciousness, feelings, personal values, without an understanding of the world – other than that given by humans based on algorithms – and without having a physicality and that relationship between mind and body that makes us human. The danger of these mirrors is a recurring theme in representations broadcast in the public space, but fundamentally, it is not the machines that are a danger, but we, humans, are 'lost in reflection' (Vallor, 2024, p. 34). When the replicant Roy Batty, in the 1982 Ridley Scott film *Blade Runner*, says 'I've seen things you people wouldn't believe. Attack ships on fire off the shoulder of Orion. I watched C-beams glitter in the dark near the Tannhäuser Gate. All those moments will be lost in time, like tears in rain. Time to die', he does not speak about an artifact driven by a set of algorithms, but about the human condition. When we look at it, we do not see a machine, we see ourselves, the humans.

I know people who expect AI to help them overcome their human condition and transfer their brains into a computer. This desire says little about technology, and much about human life and death. It is true that there are disturbing experiments on the border of AI and biology (Farina, 2022, p. 23), where artificial intelligence systems are associated with or based on living cells, paving the way for hybrid organisms. When AI could acquire a body and become AGI, questions arise that are hard to answer: could it come to have self-awareness?, do we want this?. A moratorium on the development of AGI (Yampolskiy, 2022, p. 243) similar to the one related to the prohibition of human cloning can even be discussed.

Human condition is a subject debated for thousands of years, and all we have is a continuous and endless search. What does it mean to be human and how will humanity evolve? What will the relationship be with another intelligence, possibly one that humans have created? Technology (or ‘τέχνη’ to be more precise) intertwines with philosophy and other social-human sciences, and I would dare to say even with theology (Benek, 2020) in an indivisible way. In short, philosophy strikes back.

In a key scene from the film *A.I.*, directed by Steven Spielberg, people destroy machines in a stadium. They cheer the torture and killing of suffering machines. But they refuse this when the artifact has the face of a child. Perhaps because people perceive that the difference between adults and children carries the difference between dead and alive (Lévi-Strauss, 1952). What we see here is nothing but a mirror of ourselves – people torture and kill other people every day. Plus, a question for a future we are building: what will the rules be if AI had a conscience and suffered?

As they permeate all aspects of our lives, one can imagine many situations in which machines would be faced with impossible decisions. The harms of losing control of AI can be extremely serious (Whittlestone and Clarke, 2024). In Stanley Kubrick’s *2001*, Hal was a supercomputer willing to kill people to protect its mission. This is why we should be wary of scenarios that envisage the transfer of control to artificial intelligence. Autonomous vehicles have been talked about for decades, and although the progress made is extraordinary, we do not see autonomous vehicles roaming the streets globally. Because the question arises: who is responsible for the wrong decisions made?

We can discuss the normative relationships between people and machines when they are involved in common cooperative actions. The owner of the artifact remains responsible for its actions, so the user can criticize the machine in the sense of asking the owner to change its actions, but the artifact also has the right – and it is debatable whether it has the obligation – to signal ‘if the human acts in ways that do not contribute to their joint activity’ (Löhr, 2022, p. 189). Let us illustrate this with a previously mentioned everyday activity: supermarkets that replace cashiers with machines. When a customer arrives there, and the artifact does not function in the sense of the common cooperative action (the mundane shopping), he would have the right to ask it to correct its actions. It is not ridiculous

to criticize the device in the sense that you expect it to be capable of cooperation, and the owner should have acted accordingly. On the other hand, the device should show you if you are doing something wrong regarding cooperation. From my limited experience, currently, the artifact does not always signal what you did wrong or if it is non-functional, and the standard solution is to go to another one. There are some people who oversee what buyers are doing (in one form or another) and intervene when necessary. There are two important aspects here: we need rules even just to establish cooperation, and from an empirical approach, we still see a lot of human action, while we are told – more or less subtly – that humans are the weak link.

We already live in a world of machines that know almost everything about who we are. But they do not know who we might become beyond statistical predictions, they do not understand the implications that the general human need for the sacred can have on a person, on any person, they do not know our power to change (Vallor, 2024, pp. 56–57), they do not have our values. We can step out of the way; we tend to get lost in this system of mirrors. Not by destroying the mirrors, we are not the Luddites, but through a multidisciplinary approach and a liberal vision that involves respect for the rights and freedoms of people, but also for coherent regulations, norms and procedures.

The European Parliament is a pioneer here, alongside other institutions of the European Union, mainly the Commission and the Committee of the Regions, through the recommendation on *Civil Law Rules on Robotics* (European Parliament, 2017). It is noteworthy how the European Parliament approaches the issue of AI in a multidisciplinary way from the beginning, through literature. Right in the first point of the introduction, this resolution recalls where we start from, ‘whereas from Mary Shelley’s *Frankenstein’s Monster* to the classical myth of *Pygmalion*, through the story of Prague’s *Golem* to the robot of Karel Čapek, who coined the word, people have fantasized about the possibility of building intelligent machines’ (p. 1). And the first general principle stated (p. 4) refers to Asimov’s Laws which ‘must be regarded as being directed at the designers, producers and operators of robots, including robots assigned with built-in autonomy and self-learning’. These laws were written in science fiction stories, and they are: ‘One, a robot may not injure a human being under any conditions – and, as a corollary, must not permit a human being to be injured because of inaction on his part. [...] Two, [...] a robot must follow all orders given by qualified human beings as long as they do not conflict with Rule 1. [...] Three: a robot must protect his own existence, as long as that does not conflict with Rules 1 and 2’ (Asimov, 1942, p. 100). The three laws were completed in the form of ‘No robot may harm a human being; nor, through inaction, may he allow one to come to harm’ (Asimov, 1950, p. 51). Later, the European Parliament clearly speaks about the need for a multidisciplinary approach: ‘cross-sectorial and multidisciplinary monitoring of robotics-based applications’ (p. 8).

### 3. Addressing the challenge in the EU

As Stix (2024) shows, the European Union has a level of coherence and coordination that is unmatched internationally regarding AI and the relationships we establish with AI. EU regulations may not serve as a globally accepted model, but they set limits and indicate what the violations of those limits are (Harding, 2024, pp. 219–220). These limits are certainly valid for those who believe that democracy is still possible under the conditions that AI has entered the game, that we can create a public space where different voices can be heard, and that human nature can make the common good possible.

There is a whole series of documents from EU institutions that establish norms and regulations. Three essential segments stand out here, given that there are numerous aspects covered by European regulations. First of all, the protection of human rights and civil liberties has been an aspect established since the founding of the Union. Respect for human rights is an essential fact, because if we do not do this, any harm is possible. And we all know how easily one can end up in a situation where elaborate explanations are provided for why we should give them up. But we cannot give them up without losing our identity. There is a profoundly political choice here. Because we must answer the question: who are we, Europeans? Regulating AI in terms of protecting human rights, preventing abuse, and ensuring responsible use is a necessity outlined in European documents. Secondly, we see a social, or rather socio-economic approach to the relationship with AI, an approach that again comes from the founding of the Union. Without this social and socio-economic component, Europe as we understand it today was not and is not possible. It refers to social rights, workers' rights, not leaving anyone behind, solidarity and cohesion funds, and to the common market<sup>2</sup> as a fundamental element of economic growth. Thirdly, human control over AI is a rule that guarantees the first two aspects and on which a whole regulatory framework is built.

I briefly recall only a few European documents with significant relevance in the field, trying to highlight some of the specific ideas that shape the European approach, noting that two stages can be distinguished in outlining EU's position: the first stage, in the years 2017–2019, when the legislative-regulatory foundation is laid down, upon which the specific approach to AI is further built, and the second in 2024–2025, when the EU becomes more ambitious, aiming to become a leader in AI, and, like in a symphony, revisits and details the stated themes, eventually proposing a consistent model that can serve as a solid basis for global discussion.

I have already mentioned the European Parliament's resolution on *Civil Law Rules on Robotics* (2017) and the multidisciplinary approach. It highlights the need for ethical

---

2 The common market remains, even today, a goal, although we have taken steps in this direction, they are far from being sufficient. In a discussion with companies in the field, it was pointed out that the American market is a unique market that is immediately accessible, while the European market is not.

principles – ‘establishing basic ethical principles to be respected in the development, programming and use of robots and AI and in the incorporation of such principles into Union regulations and codes of conduct’ (p. 4), raises the issue of ‘legal liability for damage caused by a robot’ (p. 5) and advocates for a careful and gradual approach and constant monitoring of the effects. This prudence is justified (Yampolskiy, 2022, p. 238) because we know from the history of digital evolution that regulations lag behind hacking, viruses, etc.

We then have the Communication from the Commission, *Artificial Intelligence for Europe* (EC, 2018), practically a strategy aimed at making the EU ‘the champion of an approach to AI that benefits people and society as a whole’ (p. 2) and which believes that ‘we can place the power of AI at the service of human progress’ (p. 19). Here, fundamental values are present. We discussed earlier that we do not live in the best of all possible worlds. However, if we do not fight for these values, it is very likely that we will live in the worst of all possible worlds.

Essentially, the Commission’s Communication *Building Trust in Human-Centric Artificial Intelligence* (EC, 2019) speaks about human centricity and human control in the use of AI. It shows that ‘trust is a prerequisite to ensure a human-centric approach to AI’ and that ‘trustworthiness of AI should be ensured’ (pp. 1–2). Trust is an essential response to the fear of AI. On one hand, there is experts’ fear, and here the document continues the line of multidisciplinary approach and emphasizes as an important conclusion the need for ‘close cooperation of all concerned stakeholders’ (p. 10). On the other hand, there is public fear. There is a tendency to anthropomorphize AI among the public (Zhang, 2024), and the marketing of AI and its presence in pop culture reinforce this anthropomorphization, which also brings fear of machine dominance.

The year 2024 was the year in which the European Union developed its contemporary approach, based on the same values and principles already stated, but going further regarding the ambition for results and impact in real life and in the regulations that support and regulate these ambitions. The essential documents of the year 2024 are: Commission’s decision *Establishing the European Artificial Intelligence Office* (EC, 2024a); Commission’s communication *On Boosting Startups and Innovation in Trustworthy Artificial Intelligence* (EC, 2024b), which begins by discussing ‘a strategic investment framework in trustworthy AI’ and shows the EU’s ambition to be an economic leader in the field of AI, but in a regulated area; *Artificial Intelligence Act* (Regulation (EU) 2024/1689), which represents a comprehensive system of legal regulations for AI with a human-centered approach and careful attention to the voices of all stakeholders; respectively, the Opinion of the European Committee of the Regions *Ethical Artificial Intelligence and Access to Supercomputing for Start-ups* (CoR, 2024). The document from the Committee of the Regions has the particularity of highlighting the need for the involvement of regional and local authorities in the governance system of AI, the necessity of using AI in the public sector to improve the everyday lives of people. It is backed by a liberal political vision, based on norms, respect for rights, and the need for economic development, raising the need for public policies to

create ‘a skilled workforce capable of both advancing and working with AI and supercomputing technologies’ (p. 6), an aspect that the European Commission will develop in 2025 and 2026.

By the end of 2024, we had a set of regulations and standards with a multidisciplinary approach to AI, based on fundamental values and building tools to ensure the implementation of the European vision and a governance architecture. The European AI Office operates within the Commission and is responsible for developing the EU’s expertise and capabilities in the field and ensuring compliance with the established rules (as we also have established sanctions). The Artificial Intelligence Board includes high-level representatives from each state and provides advice and assistance to the Commission and Member States in the field. The Advisory Forum includes the voices of stakeholders and ensures that they are heard. The Scientific Panel is a group of independent experts, a type of institution that ensures things do not get out of hand. Moreover, the issue of safety has been addressed, we had defined categories of risks, we established what an unacceptable risk means (such as assigning a social score for public and private purposes, biometric classification of individuals to deduce or assume race, political opinions, union membership, religious or philosophical beliefs, or sexual orientation, police activities based on predictive analysis targeting individuals, etc.). Last but not least, we committed to supporting major investments in artificial intelligence through European funding programs Horizon Europe and Digital Europe.

In 2025, the European Commission introduced two new commitments with consistent impact. *The AI Continent Action Plan* (EC, 2025a) demonstrates the ambition and determination of the Union to be a leader not only in the field of rules and regulations but also in specific areas with the potential for major economic impact. It discusses companies’ access to supercomputers (p. 4), access to facilities for developing and training complex AI models (p. 9), Digital Innovation Hubs (p. 14) with applications for the industrial sector, the public sector, scientific research and about ‘upskilling and reskilling’ (p. 20) the workforce.

*The Union of Skills* document talks about the ‘transformation speed gap’ (EC, 2025b, p. 3) and shows that businesses are struggling to find workers, not enough people are gaining the skills they need, educational systems are not keeping pace with the needs of the real economy, and AI technologies are not being used equally across European countries, leading us to the old problem of brain drain. These issues are already well-known, prominently raised in the *Opinion of the European Committee of the Regions* from 2024, as well as in the reports of Enrico Letta (2024) and Mario Draghi (Draghi *et al.*, 2024a; 2024b). The proposed solution is to deepen the Union, in the sense of a common market, common education, increasing competitiveness, and adapting people and institutions to what some call digital transformation, the digital revolution, or even ‘the rise of the machines’. It is also an attempt to restore the social contract (which today seems to be at risk) and the social revaluation of work, because we need a construction of economic and social institutions (Korinek and Juelfs, 2024) to formulate coherent responses to the changes brought about by AI.

The Committee of the Regions believes that the public sector at local and regional levels should be more involved in how the relationship with AI is built in the Union. Being the authorities that work directly with citizens, this is essential for monitoring what happens in real life. In a remarkable analysis of the impact of AI on public organizations, Bullock *et al.* (2024), drawing from the classical models of Max Weber and Herbert Simon, show that these technologies have already led to an evolution in many organizations, and the changes raise serious issues concerning the ‘dehumanization of the decision-making process’ and ‘human control’, raising questions about public goods, the common good and individual freedom.

There is another discussion that we will need to have at some point, at the national level, regarding what will happen after we surpass the phase of minimal adaptation to digital transformations, after relatively simple automations and the connections between databases are implemented to provide quality and quick services to citizens. The discussion is related to a potential barrier to the expansion of AI in the public sector, beyond the classic resistance to change and beyond the limited resources that are allocated in many cases. Learning is a fundamental process to become human. From what I understand, it is also needed to become a machine. And unlike some people, machines are always learning. When learning, mistakes are inevitable; they are actually part of the learning process. If you don’t notice mistakes, you are either a genius of algorithms (sets of mathematical operations that process training data and then test data), or something isn’t working (Joshi, 2024, p. 130). This situation concerning the occurrence of errors happens, obviously, also with humans. However, public servants (in the broad sense) are dreading mistakes (which are inevitable but formally unacceptable), no matter how small. This leads to a rejection of everything the public servant considers to be out of his/her control and/or that could lead to any errors. In addition, public officials are and must be responsible for the decisions they make. It is not clear how this happens if they rely on AI for their decisions. If the machine makes a mistake, do we establish that the manufacturer is to blame? Or do we simply accept the rule of unpredictable systems/opaque algorithms and resign ourselves to the fact that the machine continues to make errors – leaving us with a national regulatory framework that is outdated and ill-suited to current realities?

#### **4. Conclusions**

Who is afraid of AI? It would be desirable for all of us who wish for a democratic AI culture to have a well-tempered fear, all of us who believe that ‘man is the measure of all things’, as Protagoras said. We have always pushed the boundaries of computing power and its applications further. Changes are already here, and they will not stop. And with them, the perception of what AI is, or should be, has evolved. I am referring to a perception related to the symbols circulated in the public space: we always want more. We make jumps that put pressure on socio-economic systems and raise new problems, without

clarifying the existing ones. There are also trends of industry concentration in a few companies that know (almost) everything about us. Companies that work very well with power. And beyond the questions related to how power uses control tools – and it does it very well already, the security of these tools is constantly under increasing attack.

Speaking about AI, we are talking about the human condition, about the relationship between creator and creature, about values, and about power. AI is not a universal panacea for our problems, which remain the same, but rather I believe we have reached the point where humankind plays the role of the sorcerer's apprentice. We all know Goethe's poem, which reprises an ancient myth, where the sorcerer's apprentice unleashes powers that he cannot control. Looking to the future, skepticism is always a solid option, but human creativity can push things far beyond imagination. Pessimistic scenarios are not inevitable. AI reflects humanity, reflecting the best or the worst in us (Harding, 2024, p. 11). We determine what the true propositions are and what the courses of action are. It is our decision to choose between good and evil.

How do we respond to fear and the entire AI issue? Globally, we already have a first response that seeks to place AI within the frameworks of ethical and democratic political norms. We have a European response. We have a set of rules, institutions, and processes that will coordinate the actions and interactions of the involved actors, including AI developers, to establish responsibilities and achieve desirable outcomes for both them and the European community. It is a liberal approach based on rules and procedures, as well as values such as respect for human rights and the assurance of civil rights and freedoms.

It is also a recognition of the necessity for a multidisciplinary approach, as the boldest questions posed by philosophers can find answers in what engineers and programmers can provide (Buckner, 2024, pp. 345–347), and, on the other hand, philosophers, theologians, and thinkers from the socio-human area are called to establish the norms that shape AI. The human condition has been explored by philosophers and creators throughout history, and their works are perfectly relevant today because they get to the essence of things.

The European response is desirable to be refined, monitored, and constantly questioned if it takes into account fundamental values; mistakes will have to be recognized and corrected. I have noticed that a constant subject addressed is education, which is always an excellent response, acknowledging the fact that the established formal educational system is being left behind by the speed of change. There is also a need for holistic training that will create technology experts (I use this term in a broad sense) who can understand philosophical aspects and can engage in dialogue and collaborate with philosophers who understand technical aspects and with other experts who have a broad openness beyond their specialization.

And then again: should we fear AI? Only if, as Asimov says (1950, p. 68), 'Mankind has lost its own say in its future'.

## References:

1. Agerwala, T., 'Artificial Intelligence: A Case for Ethical Design and Multidisciplinary', in Menon, S., Todariya, S. and Agerwala, T. (eds.), *AI, Consciousness and The New Humanism. Fundamental Reflections on Minds and Machines*, Singapore: Springer, 2024, pp. 55–74.
2. Asimov, I., 'The Evitable Conflict', in *Astounding Science-Fiction*, vol. XLV, no. 4, New York: Street & Smith Publications, June, 1950, pp. 48–68.
3. Asimov, I., 'Runaround', in *Astounding Science-Fiction*, New York: Street & Smith Publications, vol. XXIX, no. 1, March, 1942, pp. 94–104.
4. Assaad, Z., 'ChatGPT Just Passed the Turing Test. But That Doesn't Mean AI Is Now as Smart as Humans', *The Conversation*, Europe Edition, April 8, 2025, [Online] available at <https://the-conversation.com/chatgpt-just-passed-the-turing-test-but-that-doesnt-mean-ai-is-now-as-smart-as-humans-253946>, accessed on April 12, 2025.
5. Benek, C.J., 'Three Theologies That Influence How We View AI, Technology, and the World', in Hrynkow, C. (ed.), *Spiritualities, Ethics, and Implications of Human Enhancement and Artificial Intelligence*, Wilmington: Vernon Press, 2020, pp. 209–226.
6. Buckner, C.J., *From Deep Learning to Rational Machines: What the History of Philosophy Can Teach Us about the Future of Artificial Intelligence*, New York: Oxford University Press, 2024.
7. Bullock, J.B., Huang, H., Kim, K.C. and Young, M.M., 'The Challenge of AI Governance for Public Organizations', in Bullock, J.B., Chen, Y.C., Himmelreich, J., Hudson, V.M., Korinek, A., Young, M.M. and Zhang, B. (eds.), *The Oxford Handbook of AI Governance*, New York: Oxford University Press, 2024, pp. 198–209.
8. Coeckelbergh, M., *The Political Philosophy of AI*, Cambridge: Polity Press, 2022.
9. Crockett, L.J., *The Turing Test and the Frame Problem. AI's Mistaken Understanding of Intelligence*, Norwood: Ablex Publishing Corporation, 1994.
10. Dafoe, A., 'AI Governance: Overview and Theoretical Lenses', in Bullock, J.B., Chen, Y.C., Himmelreich, J., Hudson, V.M., Korinek, A., Young, M.M. and Zhang, B. (eds.), *The Oxford Handbook of AI Governance*, New York: Oxford University Press, 2024, pp. 21–44.
11. Danesi, M., *AI-Generated Popular Culture. A Semiotic Perspective*, Cham: Springer, 2024.
12. de Sio, F.S., *Human Freedom in the Age of AI*, New York & Abingdon: Routledge, 2024.
13. Draghi, M. *et al.*, 'The Future of European Competitiveness. Part A – A Competitiveness Strategy for Europe', Luxembourg: Publications Office of the European Union, September 2024a.
14. Draghi, M. *et al.*, 'The Future of European Competitiveness. Part B – In-depth Analysis and Recommendations', Luxembourg: Publications Office of the European Union, September 2024b.
15. Dreyfus, H.L., *What Computers Can't Do: A Critique of Artificial Reason*, New York: Harper & Row, 1972.
16. Ellul, J., *La technique ou l'enjeu du siècle*, Paris: Economica, 2008.
17. European Commission, 'Commission Decision of 24.1.2024 establishing the European Artificial Intelligence Office', C(2024) 390 final, Brussels, 24.01.2024a, published in the Official Journal of the European Union, C/2024/1459 from 14.02.2024.
18. European Commission, 'Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions – Artificial Intelligence for Europe', COM(2018) 237 final, Brussels, 25.04.2018.

19. European Commission, ‘Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – Building Trust in Human-Centric Artificial Intelligence’, COM(2019) 168 final, Brussels, 8.04.2019.
20. European Commission, ‘Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on Boosting Startups and Innovation in Trustworthy Artificial Intelligence’, COM(2024) 28 final, Brussels, 24.01.2024b.
21. European Commission, ‘Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – AI Continent Action Plan’, COM(2025) 165 final, Brussels, 9.04.2025a.
22. European Commission, ‘Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions – The Union of Skills’, COM(2025) 90 final, Brussels, 5.03.2025b.
23. European Committee of the Regions, ‘Opinion of the European Committee of the Regions – Ethical Artificial Intelligence and Access to Supercomputing for Start-ups’, Rapporteur: Emil Boc, published in the Official Journal of the European Union, C/2024/7066 from 4.12.2024.
24. 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), published in the Official Journal of the European Union, L 2024/1689 from 12.07.2024.
25. European Parliament, ‘European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics’, 2015/2103(INL), published in the Official Journal of the European Union, C 252 from 18.07.2018.
26. Farina, L., ‘Artificial Intelligence Systems, Responsibility and Agential Self-Awareness’, in Müller, V.C. (ed.), *Philosophy and Theory of Artificial Intelligence 2021*, Cham: Springer, 2022, pp. 15–25.
27. Foucault, M., *Surveiller et punir. Naissance de la prison*, Paris: Gallimard, 1975.
28. Garfinkel, B., ‘The Impact of Artificial Intelligence: A Historical Perspective’, in Bullock, J.B., Chen, Y.C., Himmelreich, J., Hudson, V.M., Korinek, A., Young, M.M. and Zhang, B. (eds.), *The Oxford Handbook of AI Governance*, New York: Oxford University Press, 2024, pp. 86–108.
29. Gates, B., ‘The Future of Agents. AI Is about to Completely Change How You Use Computers’, *Gates Notes*, November 9, 2023, [Online] available at <https://www.gatesnotes.com/AI-agents>, accessed on March 22, 2025.
30. Häggström, O., ‘Artificial General Intelligence and the Common Sense Argument’, in Müller, V.C. (ed.), *Philosophy and Theory of Artificial Intelligence 2021*, Cham: Springer, 2022, pp. 155–160.
31. Harding, V., *AI Needs You: How We Can Change AI’s Future and Save Our Own*, Princeton & Oxford: Princeton University Press, 2024.
32. Heidegger, M., *The Question Concerning Technology, and Other Essays*, New York & London: Harper Torchbooks, 1977.

33. Joshi, A.V., *Artificial Intelligence and Human Evolution: Contextualizing AI in Human History*, New York: Apress Media, 2024.
34. Korinek, A. and Stiglitz, J.E., 'Artificial Intelligence and Its Implications for Income Distribution and Unemployment', Working Paper 24174, December 2017, Cambridge, Massachusetts: National Bureau of Economic Research.
35. Korinek, A. and Juelfs, M., 'Preparing for the (Non-Existent?) Future of Work', in Bullock, J.B., Chen, Y.C., Himmelreich, J., Hudson, V.M., Korinek, A., Young, M.M. and Zhang, B. (eds.), *The Oxford Handbook of AI Governance*, New York: Oxford University Press, 2024, pp. 746–776.
36. Lazar, S., 'Power and AI: Nature and Justification', in Bullock, J.B., Chen, Y.C., Himmelreich, J., Hudson, V.M., Korinek, A., Young, M.M. and Zhang, B. (eds.), *The Oxford Handbook of AI Governance*, New York: Oxford University Press, 2024, pp. 198–209.
37. Letta, E., 'Much More Than a Market – Speed, Security, Solidarity. Empowering the Single Market to Deliver a Sustainable Future and Prosperity for all EU Citizens', report presented to the European Council, April 2024.
38. Lévi-Strauss, C., 'Le Père Noël supplicé', in *Les Temps Modernes*, no. 77, Paris: Les Éditions Gallimard, 1952, pp. 1572–1590.
39. Löhr, G., 'Robot Rights in Joint Action', in Müller, V.C. (ed.), *Philosophy and Theory of Artificial Intelligence 2021*, Cham: Springer, 2022, pp. 179–191.
40. O'Shea, L., *Future Histories: What Ada Lovelace, Tom Paine, and the Paris Commune Can Teach Us about Digital Technology*, London & New York: Verso, 2019.
41. Parry, R., 'Episteme and Techne', in Zalta, E.N. and Nodelman, U. (eds.), *Stanford Encyclopedia of Philosophy*, Winter 2024 edition, [Online] available at <https://plato.stanford.edu/entries/episteme-techne/>, accessed on January 18, 2025.
42. Romele, A., *Digital Habitus. A Critique of the Imaginaries of Artificial Intelligence*, New York & Abingdon: Routledge, 2024.
43. Searle, J., 'Minds, Brains, and Programs', 1980, *Behavioral and Brain Sciences*, vol. 3, no. 3, pp. 417–457, DOI: 10.1017/S0140525X00005756.
44. Stix, C., 'The Ghost of AI Governance Past, Present, and Future: AI Governance in the European Union', in Bullock, J.B., Chen, Y.C., Himmelreich, J., Hudson, V.M., Korinek, A., Young, M.M. and Zhang, B. (eds.), *The Oxford Handbook of AI Governance*, New York: Oxford University Press, 2024, pp. 937–958.
45. Susser, D., 'Artificial Intelligence and the Body: Dreyfus, Bickhard, and the Future of AI', in Müller, V.C. (ed.), *Philosophy and Theory of Artificial Intelligence*, Berlin & Heidelberg: Springer, 2013, pp. 277–287.
46. Turing, A., 'Intelligent Machinery, A Heretical Theory (c. 1951)', in Copeland, B.J. (ed.), *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life: Plus The Secrets of Enigma*, New York: Oxford University Press, 2004, pp. 465–475.
47. Vallor, S., *The AI Mirror. How to Reclaim Our Humanity in an Age of Machine Thinking*, New York: Oxford University Press, 2024.
48. Weissinger, L.B., 'AI, Complexity, and Regulation', in Bullock, J.B., Chen, Y.C., Himmelreich, J., Hudson, V.M., Korinek, A., Young, M.M. and Zhang, B. (eds.), *The Oxford Handbook of AI Governance*, New York: Oxford University Press, 2024, pp. 619–638.

49. Whittlestone, J. and Clarke S., 'AI Challenges for Society and Ethics', in Bullock, J.B., Chen, Y.C., Himmelreich, J., Hudson, V.M., Korinek, A., Young, M.M. and Zhang, B. (eds.), *The Oxford Handbook of AI Governance*, New York: Oxford University Press, 2024, pp. 45–64.
50. Yampolskiy, R.V., 'AI Risk Skepticism', in Müller, V.C. (ed.), *Philosophy and Theory of Artificial Intelligence 2021*, Cham: Springer, 2022, pp. 225–248.
51. Zhang, B., 'Public Opinion toward Artificial Intelligence', in Bullock, J.B., Chen, Y.C., Himmelreich, J., Hudson, V.M., Korinek, A., Young, M.M. and Zhang, B. (eds.), *The Oxford Handbook of AI Governance*, New York: Oxford University Press, 2024. pp. 553–571.