



ARTIFICIAL INTELLIGENCE (AI) AS AN EXISTENTIAL THREAT: THE ROLE OF THE EUROPEAN UNION AND ITS AI ACT IN TRANSNATIONAL POLICYMAKING

Uroš Čemalović, PhD, Senior Research Associate

Institute of European Studies

Trg Nikole Pašića 11, 11000 Belgrade, Serbia

contact@uroscemalovic.com

ABSTRACT

Apart from various benefits artificial intelligence (AI) brings for humanity, recent years have generated greater awareness of multifaceted dangers and threats it could represent. Among them, the most usually invoked are the disturbance of the labor market, potential threats for the human creativity and protection of intellectual property rights, human alienation, risks of manipulative media content and cross-border cyber interference, and, finally, the greatest threat of all: the existential one. Already well-known and thoroughly discussed potential causes of catastrophic outcomes for humanity include all-encompassing threats, such as climate change, nuclear war, or pandemics. While terminology varies, threats that could result in partial or complete destruction of humanity are typically termed “existential risk/threat” or a threat to “existential security”. On the list of possible threats, one of the most recent entries is the danger to humanity’s survival posed by the actions of some advanced, potentially self-aware, and highly destructive AI.

The EU Regulation 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence (EU AI Act) is the first supranational legally binding act of this kind. However, due to various reasons – such as the EU’s internal political dynamics and the limitations imposed by its decision-making processes – the EU AI Act is the smallest common denominator of its Member States’ regulatory intentions, leading to controversial, complex and often unenforceable legal norms. While Art. 5 of the EU AI Act is dedicated to different prohibited AI practices, the EU legislation does not offer an applicable regulatory framework dedicated to existential threat from AI. At the same time, different trans-national initiatives – such as Center for AI Safety, Future of Life Institute, AI Red Lines – have already proposed different policy and regulatory measures to mitigate the risk of extinction from AI.

The article first focuses on the notion of existential threat from AI and the potential self-awareness AI could achieve (Chapter 2), before turning to the issue of global regulatory initiatives aiming to mitigate the most significant AI risks (Chapter 3); then it examines the existing EU legislation, as well as the Union's current and potential role in transnational AI-related policymaking (Chapter 4). The methodological approach is based on dynamic and holistic legal content analysis of the EU legislation and policy documents, while the literature review is combined with references to relevant jurisprudence and insights into legal reasoning of policy makers and other relevant stakeholders. Finally, whenever appropriate, the analysis is based on convergence of legal studies with interdisciplinary explorations of political science, psychology, and technology.

Keywords: Artificial intelligence, EU AI Act, European Union, Existential threat, Global regulation of AI

1. INTRODUCTION

As counter-intuitive as it may sound, the astonishing progress of the science in the last several decades has not made shorter the list of all-encompassing, global threats for humanity. Apart from deadly weather perils and pandemics, which have been decimating the humans in numerous occasions in the previous centuries,¹ the technological progress added some new, previously totally unknown threats, such as nuclear war, but also augmented risks related to human-driven climate changes.² In the current geopolitical context, the humanity's "failure to establish existential security"³ is particularly striking, while the list of potential existential threats has become even longer. The newly identified threat is artificial intelligence (AI), while more and more organizations and scientific experts in various fields continue to warn about the danger it could pose. Despite the numerous benefits AI offers humanity, recent years have heightened the awareness of its multifaceted risks. Among these, the most commonly cited are labor market disruptions, threats to human creativity and intellectual property protection, human alienation, risks from manipulative media content and cross-border cyber interference, and above all, the existential threat. The European Union was one of the frontrunners in adopting legislation dedicated to the regulation of AI, but its regulatory response has been excessively burdened by long decision-making processes and limited scope of legislative intervention. The article will, therefore, first critically examine the notion of the existential threat from AI and how it is intertwined

¹ For a detailed overview, see Piret, J.; Boivin, G., *Pandemics Throughout History*, Frontiers in Microbiology Vol. 11, 2021, pp. 1-16.

² For different, but equally valuable approaches, see McLaughlin, A., *Existential Risk, Climate Change, and Nonideal Justice*, The Monist, Vol. 107, No. 2, 2024, pp. 190-206; Huggel C. *et al.*, *The existential risk space of climate change*, Climatic Change Vol. 174, No. 1-2, 2022, pp. 1-20.

³ Lewis-Jones, R., *Addressing Humanity's failure to establish Existential Security*, Journal of International Political Theory, Vol. 22, No. 1, pp. 92-110. See also Scouras, J., *Nuclear War as a Global Catastrophic Risk*, Journal of Benefit-Cost Analysis, Vol. 10, 2019, pp. 274-295.

with its potential self-awareness (Chapter 2), before turning to the issue of global regulatory initiatives to mitigate major AI risks (Chapter 3); the focus will then be on the existing EU legislation applicable to the regulation of AI, as well as the Union's current and potential role in transnational AI policymaking (Chapter 4).

2. AI AS AN EXISTENTIAL THREAT

In various doctrinal studies and global initiatives dedicated to different potential dangers AI could represent, diverging notions and approaches have been used. Whereas some sources mention the threat to “existential security”⁴ AI could represent, or evoke the eerie perspective of “existential anxiety”⁵ it provokes, others mention “existential risk”⁶, seen as “the hypothetical threat posed by an artificial system that would possess both the capability and the objective, either directly or indirectly, to eradicate humanity”.⁷ While it is not only difficult, but also unnecessary to make a clear distinction between the notions of existential risk/threat, or threat to existential security, in order to illustrate how the already existing risks from AI can slowly transform in existential threat, it seems perfectly appropriate to quote one particularly strong part of an open letter, titled “Pause Giant AI Experiments”, published by the Future of Life Institute on 22 March 2023: “Should we let machines flood our information channels with propaganda and untruth? Should we automate away all the jobs, including the fulfilling ones? Should we develop nonhuman minds that might eventually outnumber, outsmart, obsolete and replace us?”.⁸ However, if it goes beyond any doubt that AI-driven deepfakes are more and more present in media, while automated AI-based systems continue to replace human workers, these dangers cannot be defined as an existential threat taken *stricto sensu*. On the other hand, the perspective in which “nonhuman minds” are outnumbering and replacing humans already represents a scenario endangering the very existence of mankind.

⁴ For some sources using this terminology, see Subasri, V. *et al.*, *Potential for near-term AI risks to evolve into existential threats in healthcare*, BMJ Health & Care Informatics, Vol. 27, 2025; Bernstein, S. *Existential security and the governance challenge: Confronting the antinomies of securitisation*, Global Policy Vol. 14, 2023, pp. 638-642.

⁵ Alkhalifah, J.M. *et al.*, *Existential anxiety about artificial intelligence (AI) – is it the end of humanity era or a new chapter in the human revolution: questionnaire-based observational study*, Frontiers in Psychiatry, Vol. 15, 2024, pp. 1-13.

⁶ Hoes, E.; Gilardi, F., *Existential risk narratives about AI do not distract from its immediate harms*, Proceedings of the National Academy of Sciences of the USA Vol. 122, No. 16, 2025, e2419055122, pp. 1-18.

⁷ VanRullen, R., *AI Consciousness and Existential Risk*, Available at: [arXiv:2511.19115v1] [cs.AI], 2025, Accessed 7 February 2026.

⁸ Future of Life Institute, *Pause Giant AI Experiments: An Open Letter*, Available at: [https://futureoflife.org/open-letter/pause-giant-ai-experiments/], Accessed 4 February 2026.

The question whether and, if yes, to what extent, AI could represent an existential threat for humans should begin with possible reasons for such an outcome. Although numerous scenarios can be imagined, the following two possibilities seem most likely. First, AI tools and/or systems could be, in a manner similar to the nuclear weapons, in the hands of a state, organization, or a group of persons intending to use it in order to inflict harm or as a deterrence against threats. Second, AI could – either because it has achieved self-consciousness and perceives humans as a threat, or for some other reason – independently seek to bring about the destruction of humanity or certain groups of individuals. While the second scenario remains a mere possibility, the first is already a reality. Consequently, if the challenges AI raises in the economy, media, and artistic creation have left someone indifferent to the need for a better and more detailed legal regulation of its development and use, then the perspective of the destruction of mankind should substantially change their views. In any case, the existential threat AI poses for humanity is closely intertwined with its (potential) aptitude to reach self-consciousness. This standpoint is particularly well developed by VanRullen, who claims that “consciousness could be viewed as a means towards AI alignment, thereby lowering existential risk; or, it could be a precondition for reaching certain capabilities or levels of intelligence, and thus positively related to existential risk”.⁹ In both scenarios, the linkage between the consciousness of AI systems and existential risk is very difficult to deny, and, therefore, merits more thorough examination.

Every attempt to unequivocally determine whether AI can become self-conscious – and, in the case of a positive answer, to (at least roughly) predict when this will occur – is significantly hampered by the fact that not only the human consciousness itself has not been sufficiently explored, but also that there are various concepts in its definition. Therefore, it is of a particular importance to establish certain boundaries of insight that are possible to reach. The nature of consciousness is multidimensional, so before considering AI's aptitude to possess it, it is necessary to specify which dimension and/or form of consciousness is being targeted. As Farisco and others have rightfully pointed out, “whilst the extension of other mental terms like attention and intelligence to non-human animals and AI systems may appear less problematic than consciousness, we should distinguish two semantic dimensions when referring any mental term to AI: the epistemological dimension (i.e., what is the nature of the feature indicated by the term?), and the functional dimension (i.e., how does the feature indicated by the term work? What is the relevant architecture, that is the adequate combination of structure

⁹ VanRullen, R., *op. cit.*, note 7, p. 1.

and functionalities?)”.¹⁰ If this distinction is accepted, extending the first dimension to AI is, at its current level of development, almost impossible, while “the extension of the second dimension may be more straightforward, provided that relevant structures and functions may be replicated at the adequate scale or level of details”.¹¹ Therefore, AI could potentially have some functional elements of consciousness, but its content would differ from the same property we recognize in humans. However, these elements should be sufficient to consider AI as an existential threat. In the apocalyptic scenario in which AI turns against humans, one of the key reasons for this is a quantum leap – which some authors have also named “the moment of technological singularity”¹² – taking place when AI becomes self-conscious.

There is a significant number of authors who consider that a self-conscious AI (at least, for now) is impossible. In this respect, it is particularly interesting to confront the standpoints of different authors regarding AI tools based on large language models (LLM). For example, in their article apodictically entitled “There is no such thing as conscious artificial intelligence”, Porębski and Figura affirm that “unlike human use of language, there is no intention behind the strings of characters generated by LLMs [while] their language use is stripped of reference to meanings (those existing in the world rather than represented numerically in the vectors), making them comparable to stochastic parrots”.¹³ On the other hand, in the article “Signs of consciousness in AI: Can GPT-3 tell how smart it really is?” – which, adding some awkwardness in the entire debate, was published in the same journal just a few months earlier – Bojić, Stojković and Jolić Marjanović uphold that “although there are no definitive standards and conclusions regarding its consciousness, its [GPT-3’s] ability to receive inputs (similar to reading), reason, analyze, generate predictions, and perform NLP tasks suggests some aspects of subjectivity, perception, and cognition [while] these observations highlight the need to continually examine the potential emergence of awareness and self-awareness in large language models.”¹⁴ What seems particularly important to underline is

¹⁰ Farisco, M. *et al.*, *Is artificial consciousness achievable? Lessons from the human brain*, Neural Networks Vol. 180, 2024, pp. 1-14.

¹¹ *Ibid.*, p. 10.

¹² Vinge, V., *The Coming Technological Singularity: How to Survive in the Post-Human Era*, Available at: [<https://ntrs.nasa.gov/api/citations/19940022856/downloads/19940022856.pdf>], Accessed 7 February 2026.

¹³ Porębski, A.; Figura, J., *There is no such thing as conscious artificial intelligence*, Humanities & Social Sciences Communications Vol. 12, 2025, pp. 1-12.

¹⁴ Bojić, L.; Stojković, I.; Jolić Marjanović, Z., *Signs of consciousness in AI: Can GPT-3 tell how smart it really is?*, Humanities & Social Sciences Communications Vol. 11, 2024, pp. 1-15, 8.

exactly this need to permanently examine and evaluate AI's aptitude for consciousness/awareness, chiefly because AI models are evolving with remarkable rapidity.

Finally, there is yet another important legal issue related to the status of AI: its potential personhood and legal conditions under which it could be granted. The issue is highly divisive, as well as ethically and politically sensitive. Some authors have identified new constructs to complement the taxonomy of arguments for AI legal personality, such as "cognitive strain, consciousness, cyborg paradox, reasoning replicability, relativism, AI misuse, excessive human effort and substitution".¹⁵ On the other hand, AI can become a legal subject, "namely a derivative legal subject like a legal organization,"¹⁶ but specific conditions in different legal orders significantly vary and often lack precision. The most sensitive issue regarding AI's legal personhood still remains possible consequences of such regulatory choice, especially given the absence of detailed and universally accepted guidelines on the use of AI tools.

3. GLOBAL REGULATORY INITIATIVES TO MITIGATE MAJOR AI RISKS

The number of national and supranational organizations, fora and initiatives aiming to achieve certain level of transnational regulation of major AI-related risks is constantly increasing over the last few years. In May 2023, the Center for AI Safety presented its "Statement on AI Risk", supported by a large number of influential individuals and still open for signatures. The statement consists of a single sentence: "Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war".¹⁷ What the Statement refers to as "risk of extinction" other organizations, such as the Future of Life Institute – in an article authored by Ariel Conn and published in November 2015 – term "existential risk", defined as "any risk that has the potential to eliminate all of humanity or, at the very least, kill large swaths of the global population, leaving the survivors without sufficient means to rebuild society to current

¹⁵ Staszkiwicz, P. *et al.*, *Artificial intelligence legal personality and accountability: auditors' accounts of capabilities and challenges for instrument boundary*, *Meditari Accountancy Research*, Vol. 32, No. 7, 2024, pp. 120-146.

¹⁶ Putriyanti, E.D. *et al.*, *Artificial Intelligence: Legal Status and Development in the Establishment of Regulatory*, in: Umiyati, M. *et al.* (eds.), *Proceedings of the International Conference on "Changing of Law: Business Law, Local Wisdom and Tourism Industry"* (ICCLB 2023), *Advances in Social Science, Education and Humanities Research* 804, pp. 446-457, 450.

¹⁷ Center for AI Safety, *Statement on AI Risk*, Available at: [<https://safe.ai/work/statement-on-ai-risk>], Accessed 9 February 2026.

standards of living”.¹⁸ Regardless of the term chosen, it is clear that this concerns a threat whose deepest essence lies precisely in the unpredictability of consequences arising from accelerated AI development. For this very reason, numerous initiatives have proposed and continue to propose more concrete measures to mitigate this threat, while its complete elimination appears difficult to achieve. Among the proposed measures, those particularly standing out are, on the one hand, appeals for governments to reach a binding international agreement, and, on the other, calls for a moratorium on training and developing AI models.

The initiative named “AI Red Lines”, launched in September 2025, has published a Global Call, in which it directly urges the governments of all countries to reach an international agreement on “red lines” in AI development that all parties would commit not to cross. Although the Global Call does not define precisely what it considers as red lines, this initiative is accompanied by a useful explanation. Namely, some advanced AI systems have already demonstrated the capacity for deceptive and harmful behavior, yet they are being granted ever greater autonomy to take actions and make decisions in the world, which has led many experts, “including those at the forefront of development [of AI systems], to warn that it will become increasingly difficult to exert meaningful human control in the coming years”.¹⁹ Unlike the brief statement of the Center for AI Safety, this elaborated initiative enables somewhat deeper analysis. Namely, three factors that render AI development genuinely threatening can be identified. First, the combination of the speed and unpredictability of that development, particularly given that only a small portion of information about the capabilities of advanced AI systems reaches the public. Second, some harmful effects of AI use are already known, and in combination with the previous element, it can be expected that there will be more of them, with increasingly damaging impacts. Third, the emphasis is on acting before it is too late, that is before the moment when human control becomes very difficult or impossible. Once again, the danger of the unknown, namely of AI slipping from human control and beginning to act independently, lies at the very core of the concept of existential threat from AI.

In March 2023, the Future of Life Institute published an Open Letter calling for a moratorium (lasting at least 6 months) on training and developing AI-based systems “more powerful than GPT-4”. This initiative has so far been supported by over 33,000 signatures, including internationally recognized scientists (Yuval

¹⁸ Conn, A., *Existential Risk*, Future of Life Institute, Available at: [https://futureoflife.org/existential-risk/existential-risk/], Accessed 9 February 2026.

¹⁹ AI Red Lines, *Global Call for AI Red Lines*, Available at: [https://red-lines.ai/], Accessed 9 February 2026.

Noah Harari, Max Tegmark), as well as founders or directors of major technology companies (Steve Wozniak, Elon Musk, Emad Mostaque). Alongside citing several references, the letter presents compelling argumentation, stating that contemporary AI-based systems are now becoming competitive with humans in general tasks. The Open Letter emphasizes that “powerful AI systems should be developed only once we are confident that their effects will be positive and their risks will be manageable [while] AI labs and independent experts should use this pause to jointly develop and implement a set of shared safety protocols for advanced AI design and development that are rigorously audited and overseen by independent outside experts”.²⁰ This argumentation is, undoubtedly, understandable and easy to agree with. However, core flaw in such a formulation mirrors the problem already identified in international arms race. Namely, if not all actors active in AI development accept such an initiative, it can achieve only partial or no success whatsoever, since non-accepting parties will, while others pause, gain an unattainable developmental advantage. Therefore, even individuals and companies responsible for AI development who agree with the initiative and wish to accept it often hesitate to do so, fearing they will be sidelined in the technological race.

4. THE ROLE OF THE EU AND ITS LEGISLATION IN TRANSNATIONAL AI POLICYMAKING – FOCUS ON EXISTENTIAL SECURITY

Although numerous national and international institutions, bodies and initiatives have already elaborated a wide range of recommendations, standards, guidelines and similar documents striving to contribute to the regulation of AI tools, systems and practices,²¹ the first supranational legally binding act dedicated exclusively to AI is the EU Regulation 2024/1689 of 13 June 2024 establishing harmonised rules on artificial intelligence²² (hereinafter: EU AI Act). While the existence of

²⁰ Future of Life Institute, *Pause Giant AI Experiments: An Open Letter*, Available at: [<https://futureoflife.org/open-letter/pause-giant-ai-experiments/>], Accessed 9 February 2026.

²¹ The Organization for Economic Co-operation and Development launched „The OECD.AI Policy Navigator“, defined by the Organization itself as “a living repository from more than 80 jurisdictions and organizations”, Available at: [<https://oecd.ai/en/dashboards/overview>]. However, on the international level, the repository comprises only 8 initiatives from the following organizations: the African Union, G7, the United Nations, the Association of Southeast Asian Nations, the Inter-American Development Bank, The United Nations Educational, Scientific and Cultural Organization, the EU and the OECD/GPAI itself.

²² Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) [2024] OJ L series.

significantly different national legislations in a particular area is neither prohibited nor always undesirable, in the field of AI uniform rules at the EU level have been necessary in order to make its single market viable. It is particularly worth noting that – even before the adoption of the EU AI Act, the Court of Justice of the EU (CJEU)²³ was invoking the provisions of EU legislation (such as, for example, recitals 9 and 13 of Directive 2019/1024 on open data and the re-use of public sector information²⁴) which peripherally mention AI in the context of “advanced digital technologies”.²⁵ However, given the significant number of pending cases, it is yet to be seen how the CJEU will interpret some core provisions of the EUAI Act.

In September 2024, three months after the adoption of the EU AI Act, the Council of Europe’s (CoE) Framework Convention on Artificial Intelligence and human rights, democracy and the rule of law²⁶ has been opened for signature. While the CoE has declared that this convention is “the first-ever international legally binding treaty in this field,”²⁷ legal doctrine presented more nuanced views, pointing out that “key compromises made in the negotiation process – including exemptions for private sector regulation and national security uses of AI – will likely weaken the Framework Convention’s effectiveness”.²⁸ Moreover, narrowed to the issues of human rights, democracy and rule of law, the CoE’s Framework Convention sidesteps high-risk AI systems in general, let alone the issues such as the existential threat from AI.

The European Commission published the Proposal for a Regulation of the European Parliament and of the Council establishing harmonized rules on AI in April 2021,²⁹ marking the first formal step leading to the adoption of a legally binding Union act dedicated exclusively to AI. This text was the product of lengthy and arduous discussions and negotiations, during which the two key milestones were the

²³ Case C-336/23 *HP – Hrvatska pošta d.d. v. Povjerenik za informiranje* [2024], ECLI:EU:C:2024:979.

²⁴ Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information, [2019] OJ L 172, pp. 56-83.

²⁵ Directive (EU) 2019/1024, recital 9.

²⁶ Council of Europe Treaty Series – No. 225, Available at: <https://rm.coe.int/1680afae3c>, Accessed 14 February 2026.

²⁷ Council of Europe - Framework Convention on Artificial Intelligence and human rights, democracy and the rule of law, Available at: <https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence>, Accessed 14 February 2026.

²⁸ Hoxtell, W., *Global AI Regulation at a Time of Transformation: The Council of Europe’s Framework Convention on Artificial Intelligence*, EnsuredEurope, Available at: <https://www.ensuredurope.eu/publications/global-ai-regulation>, Accessed 14 February 2026.

²⁹ For the full text of the Proposal, see: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52021PC0206>, Accessed 14 February 2026.

Commission's Communication *Artificial Intelligence for Europe*³⁰ from April 2018 and its *Report on the safety and liability implications of artificial intelligence, the Internet of Things and robotics*³¹ from February 2020. It was not until mid-March 2024, almost three years after the publication of the proposal, that the European Parliament, in the first reading under the ordinary legislative procedure, adopted the EU AI Act. Two months later, the Council of the EU adopted the same text, which was published in the Official Journal of the EU on 12 July 2024. For the sake of legal certainty and to allow sufficient time for all subjects to which the EU AI Act applies, Article 113 provides that the general rule is that the act shall apply from 2 August 2026. However, it is envisaged that certain provisions thereof shall apply from 2 February 2025 (prohibited AI practices), 2 August 2025 (e.g., provisions on general-purpose AI models), or 2 August 2027 (provisions on high-risk AI systems).

The EU AI Act did not dedicate – even distantly – a single provision to the issue of existential risks/threats AI could represent for the humanity. The main corpus of substantive provisions of the EU AI Act addresses five main themes: prohibited practices (Article 5), high-risk AI systems (Articles 6-49), transparency obligations for providers and users of certain AI systems (Article 50), and general-purpose AI models (Articles 51-63). Moreover, considering that the preamble to the EU AI Act comprises 180 recitals, praise is due to the EU legislators for the effort to justify the document's background and various regulatory solutions, indicating that the drafting process was open and democratic. However, due to the specific features of the EU's decision-making process – and above all, the slowness conditioned by the complexity of the topic and conflicting economic and political interests – the Act does not go far enough in many provisions, which might lead both to its limited effects and to significant problems in its interpretation and implementation in the EU Member States. The following analysis will first examine some major elements of the existing structure and content of the EU AI Act that could allow the future legislative steps in the direction of a more direct and straightforward regulation of AI existential risk (section 4.1), before turning to the issue of the key features of a potential supranational regulatory framework dedicated to AI and, in particular, existential threat it represents (section 4.2).

³⁰ Artificial Intelligence for Europe, COM(2018) 237 final, Available at: [<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN>], Accessed 15 February 2026.

³¹ European Commission, Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics, COM(2020) 64 final, Available at: [<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0064>], Accessed 15 February 2026.

4.1. The Pathways for the Evolution of the EU AI Act

In spite of the fact that the EU AI Act did not have any ambition whatsoever to mitigate the existential risk from AI, some of its provisions could be interpreted in a sense that, by means of the future amendments to the Act, it could represent a base – or, at least, a contextual and teleological justification – for regulating AI-related existential risk. The need for such legislative changes is beyond any doubt, while some authors have pertinently pointed out that “AI’s exponential growth outpaces static laws, rendering traditional command-and-control regulation inadequate”.³² While the EU secondary legislation has to respect the limits imposed by the treaties and, in particular, those related to the Union’s competences, the existing regulatory framework does not allow the conclusion that a common European regulation of AI existential threat is beyond reach. It is, therefore, appropriate to examine the EU AI Act’s existing provisions dedicated to the risks and harms AI may cause, as well as the issue of how its provisions on prohibited AI practices could evolve in the direction of regulating existential risks.

Recital 5 of the EU AI Acts is one of its most forward-looking provisions, stating that “depending on the circumstances regarding its specific application, use, and level of technological development, AI may generate risks and cause harm to public interests and fundamental rights that are protected by Union law”. In the context of AI-based existential threat, the key issue here is whether the survival of mankind could be considered as a “public interest” or “fundamental right” protected by EU law. Until now, this issue has been treated only partially, and in the context of environmental law.³³ However, would the Art. 2-1 of the Charter of Fundamental Rights of the EU, guaranteeing that “everyone has the right to life”, could be used a legal basis for the adoption of the amendments to the EU AI Act dedicated to AI existential threat? In the current state of the Union’s legislation and jurisprudence, there is no explicit legal base or interpretation that would indicate a positive answer to this question. Moreover, although the EU AI Act “classifies AI systems into three categories: unacceptable-risk, high-risk and limited and minimal-risk, it still falls short of the requirements needed to protect several fundamental human rights and to prevent AI from being used in ways that would aggravate existing inequities and discrimination”.³⁴ Therefore, it seems that

³² Mahmutovic, A., *The EU AI Act: a proactive framework for comprehensive AI regulation*, International Journal of Law and Information Technology, Vol. 33, 2025, pp. 1-12, 7.

³³ See Somsen, H., *From Improvement Towards Enhancement: A Regeneration of EU Environmental Law at the Dawn of the Anthropocene*, in: Brownsword, R.; Scotford, E.; Yeung K. (eds.), *The Oxford Handbook of Law, Regulation and Technology*, Oxford Handbooks, 2017.

³⁴ Federspiel, F. *et al.*, *Threats by artificial intelligence to human health and human existence*, BMJ Global Health, Vol. 8, No. 5, 2023, pp. 1-6, 3.

the first step in improving the EU AI Act would be to define the nature and scope of unacceptable risks.

While the EU AI Act comprises detailed rules related to the classification of high-risk AI systems (Art. 6), requirements related to them (Art. 8), risk management system (Art. 9), data governance (Art. 10), transparency (Art. 13), human oversight (Art. 14) and numerous other issues, there are no core provisions of the Act dedicated to unacceptable risks, an issue that has remained entirely unregulated. Apart from the two provisions of the preamble (recitals 46 and 179) mentioning *expressis verbis* the notion of “unacceptable risk”, no core provision reiterates it, let alone defines it. Moreover, the contextual interpretation of recital 26, stating that “in order to introduce a proportionate and effective set of binding rules for AI systems, a clearly defined risk-based approach should be followed [...] it is therefore necessary to prohibit certain unacceptable AI practices, to lay down requirements for high-risk AI systems,” leads to the conclusion that the category of “unacceptable risk” has only been introduced in order to justify the rules on prohibited AI practices. In other words, from the perspective of the EU legislators, AI practices demonstrating an unacceptable risk should be prohibited, while high-risk AI systems are heavily regulated. On the other hand, even a summary analysis of the EU AI Act’s provisions on prohibited AI practices reveals that they remain unrelated with AI existential threats. In conclusion, if its existing regulatory structure and logic is to be preserved, the EU AI Act could evolve towards regulating AI-based existential risks by embracing significantly wider perspective on prohibited AI practices, combined with a detailed definition of unacceptable risks.

4.2. Key Features of a Supranational Regulatory Framework

As it has already been outlined in Chapter 3, supranational regulatory initiatives propose the adoption of different measures, but substantially remain legally non-binding. On the other hand, the EU AI Act mentions different risks and regulates prohibited AI practices but remains silent when it comes to existential threat from AI. In such a context, one overarching question becomes particularly urgent: what would be the most appropriate nature, scope and content of a supranational regulatory framework dedicated to the development and use of AI systems and tools which could, on the one hand, ensure the preservation of AI’s numerous advantages while, on the other, preempt all the undeniable threats it brings? This issue becomes particularly salient in the context of the existential threat from AI, as it is impossible to imagine a greater risk that AI systems could potentially represent for humanity. The following considerations will, therefore, focus on some key characteristics of a desired and potentially most effective regulatory framework

dedicated to AI; the objective of this endeavor is also to distill how and to which extent regulatory solutions and approaches set out in the EU AI Act could serve as a starting point in adopting future normative acts with cross-border effects.

First, any transnational regulation of AI purporting to be effective must consist of legally binding norms, for which sovereign states – in the current state of development of international relations and institutions, best positioned for the task – should ensure effective methods of implementation, oversight, and sanctions. The aforementioned Global Call for AI Red Lines³⁵ also insists on this element, calling upon the governments to “reach an international agreement on red lines for AI, ensuring they are operational, with robust enforcement mechanisms” to prevent universally unacceptable risks. While in the current, politically significantly weakened system of international law, there are some conventions accepted by nearly all states, ensuring mechanisms for an effective implementation of their norms is a considerably more difficult task. All the existing international attempts to regulate the development and use of AI systems and tools are non-binding texts (initiatives, calls, recommendations, collections of good practices) which are not easy to translate into compulsory, enforceable, and effective legal norms.

Second, the very nature and fundamental characteristics of AI tools and systems impose one major feature of its successful regulation: it must be supranational. In the context of the global technological race, it is practically impossible to imagine that the results (or consequences) of the development of AI systems will not spill over from the islands where legal rules do not apply onto the territories of the states that have adopted (and enforced) a certain legal framework. As it has been pertinently remarked, “given the borderless nature of this issue, an internationally coordinated response is necessary,”³⁶ followed by a “fundamental change of mindset regarding the nature of regulation”.³⁷ Apart from non-governmental regulatory initiatives already examined in Chapter 3, various international organizations (United Nations, Organization for Security and Co-operation in Europe, G7 Group) have also drafted and published certain documents that purport to regulate the issue of AI at the global level. However, all these acts lack two key features: universal acceptance by the states and a sufficiently wide scope. In this respect, a model introduced by the EU AI Act could be a good starting point of global regulatory efforts.

³⁵ AI Red Lines, Global Call for AI Red Lines, *loc. cit.*, note 17.

³⁶ Zaidan, E.; Ibrahim, I.A., *AI Governance in a Complex and Rapidly Changing Regulatory Landscape: A Global Perspective*, Humanities and Social Sciences Communications Vol. 11, No. 1121, 2024, pp. 1-18, 9.

³⁷ Erdélyi, O.; Goldsmith, J., *Regulating artificial intelligence: Proposal for a global solution*, Government Information Quarterly, Vol. 39, No. 4, 2022, pp. 1-7, 5.

Third, the global regulation of AI should be as comprehensive as possible, covering the largest number of threats that are not only currently present, but can also be expected in the future. Critics may object that this requirement runs counter to the principles of the rule of law, legal certainty, as well as so-called protection of legitimate expectations, the latter of which is particularly well developed in the case law of the CJEU.³⁸ However, the speed of the development of AI tools and systems would render meaningless, within just a few months, any set of binding legal rules incapable of encompassing the expected harmful consequences of advanced AI systems that may emerge in the foreseeable future. The EU has, through the provisions of Article 5 of the EU AI Act on prohibited AI practices – attempted to take such a step. This regulatory effort, in numerous aspects, did not go “far enough, introducing a series of legal standards that would raise issues of interpretation, and thus potentially lead to its unsatisfactory level and quality of enforcement”.³⁹ On the other hand, the existing EU legislation – while at least offering some pathways for a potential global regulation of prohibited AI practices – should still find a genuinely European response to the issue of AI existential threat.

Fourth, given that “the traditional top-down tack to regulation may stifle innovation and creativity and stop AI from realizing its transformative potential”,⁴⁰ every desirable global, legally binding, effective, comprehensive, and forward-looking AI regulation should also be designed so as not to hinder the development of legitimate, undoubtedly positive and beneficial AI systems and tools. Given the significant obstacles to achieving any international AI regulation whatsoever, this requirement appears less significant than the previous three. Both at the international and national levels, there are so few elaborated legally binding rules on AI that it is difficult to expect any existing acts aimed at its regulation to be capable of slowing the development of AI’s positive aspects. What is urgently and critically needed is precisely the regulation of the dangers and threats that the development and use of AI can bring. In spite of its limitations, the existing legal framework introduced by the EU AI Act could represent a solid basis for the elaboration of

³⁸ There is a long-standing CJEU jurisprudence regarding this issue; see, for example, the judgements in cases C-776/23 P *Commission v Spain and Others* to C-780/23 P, ECLI:EU:C:2025:487 (ECR not yet assigned), paras 32, 39, 46, 55, 59; C-568/11 *Agroferm*, ECLI:EU:C:2013:407, para 13; C-203/07 P, *Hellenic Republic v Commission of the European Communities*, ECR 2008 I-08161, para 16; C-144/14, *Cabinet Medical Veterinar Dr. Tomoiagă Andrei v Direcția Generală Regională a Finanțelor Publice Cluj-Napoca prin Administrația Județeană a Finanțelor Publice Maramureș*, ECLI:EU:C:2015:452, paras 41, 48.

³⁹ Ćemalović, U., *Prohibited artificial intelligence practices according to Article 5 of the European Union’s regulation on AI - between the ‘too late’ and the ‘not enough’*, International Journal of Law and Information Technology Vol. 32, No. 4, 2024, pp. 1-13, 11.

⁴⁰ Unknown authors, *Co-Governance and the Future of AI Regulation*, Harvard Law Review, Volume 138, 2025, pp. 1609-1632.

a global regulatory framework for a responsible, human-centered AI, designed to eliminate (or, at least) reduce its existential threat to humanity.

5. CONCLUSION

AI-based tools and systems are developing in such a rapid pace that practically every regulatory framework aiming to regulate them can soon be rendered outdated. While numerous specific threats AI poses across diverse fields (such as, for example, its impact on the labor market, creative industries, media and protection of intellectual property rights) have already become reality – and, therefore, have received, at least partial, regulatory response – the issue of existential threat from AI is often seen as purely speculative. The critical examination of the notion of AI's existential threat and how it is intertwined with its potential self-awareness (Chapter 2) has shown that AI could – either because it has achieved self-consciousness and perceives humans as a threat, or for some other reason – seek to bring about the destruction of humanity or certain groups of individuals. The analysis of recent global regulatory initiatives to mitigate major AI risks (Chapter 3) has revealed that, among the proposed measures, those particularly standing out are, on the one hand, appeals for governments to reach a binding international agreement, and, on the other, calls for a moratorium on training and developing AI models. However, the closest reference to the existential threat from AI can be found in the “Statement on AI Risk” elaborated by the Center for AI Safety, which underlines that “mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war”. While the EU AI Act is the first supranational legally binding act dedicated exclusively to artificial intelligence, it did not dedicate a single provision to the issue of existential risks/threats AI could represent for the humanity. However, the scrutiny (Chapter 4) of its provisions on prohibited AI practices and high-risk AI systems has led to the conclusion that, in spite of its limitations, the Act could represent a solid basis for the elaboration of a legally binding, transnational and comprehensive regulatory framework, also designed to substantially mitigate AI's existential threat to humanity.

REFERENCES

BOOKS AND ARTICLES

1. Alkhalifah, J.M. *et al.*, *Existential anxiety about artificial intelligence (AI) – is it the end of humanity era or a new chapter in the human revolution: questionnaire-based observational study*, *Frontiers in Psychiatry*, Vol. 15, 2024, pp. 1-13
2. Bernstein, S. *Existential security and the governance challenge: Confronting the antinomies of securitisation*, *Global Policy* Vol. 14, 2023, pp. 638-642

3. Bojić, L.; Stojković, I.; Jolić Marjanović, Z., *Signs of consciousness in AI: Can GPT-3 tell how smart it really is?*, Humanities & Social Sciences Communications Vol. 11, 2024, pp. 1-15
4. Brownsword, R.; Scotford, E.; Yeung K. (eds.), *The Oxford Handbook of Law, Regulation and Technology*, Oxford Handbooks, Oxford, 2017
5. Ćemalović, U., *Prohibited artificial intelligence practices according to Article 5 of the European Union's regulation on AI - between the 'too late' and the 'not enough'*, International Journal of Law and Information Technology Vol. 32, No. 4, 2024., pp. 1-13
6. Erdélyi, O.; Goldsmith, J., *Regulating artificial intelligence: Proposal for a global solution*, Government Information Quarterly, 2022, Vol. 39, No. 4, pp. 1-7
7. Hoes, E.; Gilardi F., *Existential risk narratives about AI do not distract from its immediate harms*, Proceeding of the National Academy of Sciences of the USA Vol. 122, No. 16, 2025, e2419055122, pp. 1-18.
8. Huggel, C. *et al.*, *The existential risk space of climate change*, Climatic Change Vol. 174, No. 1-2, 2022, pp. 1-20
9. Farisco M. *et al.*, *Is artificial consciousness achievable? Lessons from the human brain*, Neural Networks Vol. 180, 2024, pp. 1-14
10. Federspiel, F. *et al.*, *Threats by artificial intelligence to human health and human existence*. BMJ Global Health, Vol. 8, No. 5, 2023, pp. 1-6
11. Mahmutovic, A., *The EU AI Act: a proactive framework for comprehensive AI regulation*, International Journal of Law and Information Technology, Vol. 33, 2025, pp. 1-12
12. McLaughlin, A., *Existential Risk, Climate Change, and Nonideal Justice*, The Monist, Vol. 107, No. 2, 2024, pp. 190-206
13. Lewis-Jones, R., *Addressing Humanity's failure to establish Existential Security*, Journal of International Political Theory, Vol. 22, No. 1, pp. 92-110
14. Piret, J.; Boivin G., *Pandemics Throughout History*, Frontiers in Microbiology Vol. 11, 2020
15. Porębski, A.; Figura, J., *There is no such thing as conscious artificial intelligence*, Humanities & Social Sciences Communications Vol. 12, 2025, pp. 1-12
16. Putriyanti, E.D. *et al.*, *Artificial Intelligence: Legal Status and Development in the Establishment of Regulatory*, in: Umiyati, M. *et al.* (eds.), *Proceedings of the International Conference on "Changing of Law: Business Law, Local Wisdom and Tourism Industry" (ICCLB 2023)*, Advances in Social Science, Education and Humanities Research 804, pp. 446-457
17. Scouras, J., *Nuclear War as a Global Catastrophic Risk*, Journal of Benefit-Cost Analysis, Vol. 10, 2019, pp. 274-295
18. Somsen, H., *From Improvement Towards Enhancement: A Regenesi of EU Environmental Law at the Dawn of the Anthropocene*, in: Brownsword, R; Scotford, E.; Yeung K. (eds.), *The Oxford Handbook of Law, Regulation and Technology*, Oxford Handbooks, 2017
19. Subasri, V. *et al.*, *Potential for near-term AI risks to evolve into existential threats in healthcare*, BMJ Health & Care Informatics, Vol. 27
20. Staszkiwicz, P. *et al.*, *Artificial intelligence legal personality and accountability: auditors' accounts of capabilities and challenges for instrument boundary*, Meditari Accountancy Research, Vol. 32, No. 7, 2024, pp. 120-146

21. Unknown authors, *Co-Governance and the Future of AI Regulation*, *Harvard Law Review*, Volume 138, 2025, pp. 1609-1632
22. VanRullen, R., *AI Consciousness and Existential Risk*, Available at:
23. [arXiv:2511.19115v1] [cs.AI] 2025, accessed 7 February 2026
24. Zaidan, E.; Ibrahim, I.A., *AI Governance in a Complex and Rapidly Changing Regulatory Landscape: A Global Perspective*, *Humanities and Social Sciences Communications* Vol. 11, No. 1121, 2024, pp. 1-18

COURT OF JUSTICE OF THE EUROPEAN UNION

1. Case C-203/07 P [2008] *Hellenic Republic v Commission of the European Communities*, ECLI:EU:C:2008:606
2. Case C-568/11 [2013] *Agroferm*, ECLI:EU:C:2013:407
3. Case C-144/14 [2015] *Tomoiağă Andrei v Direcția Generală Regională a Finanțelor Publice Cluj Napoca*, ECLI:EU:C:2015:452
4. Case C-336/23 [2024] *HP – Hrvatska pošta d.d. v Povjerenik za informiranje* ECLI:EU:C:2024:979
5. Case C-776/23 P to C-780/23 [2025] *European Commission, vs Kingdom of Spain et al.*, ECLI:EU:C:2025:487

EU LAW AND POLICY DOCUMENTS, ACTS OF THE COUNCIL OF EUROPE

1. Charter of Fundamental Rights of the European Union, [2012] OJ C 326
2. Council of Europe's Framework Convention on Artificial Intelligence and human rights, democracy and the rule of law, Council of Europe Treaty Series – No. 225
3. Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information, [2019] OJ L 172
4. European Commission, Artificial Intelligence for Europe, COM(2018) 237 final, Available at: [<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN>]
5. European Commission, Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics, COM(2020) 64 final, Available at: [<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0064>]
6. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), [2024] OJ L series

WEBSITE REFERENCES

1. AI Red Lines, *Global Call for AI Red Lines*, Available at: [<https://red-lines.ai/>], Accessed 9 February 2026

2. Center for AI Safety, *Statement on AI Risk*, Available at:
[<https://safe.ai/work/statement-on-ai-risk>], Accessed 9 February 2026
3. Council of Europe's, Framework Convention on Artificial Intelligence and human rights, democracy and the rule of law, Available at:
[<https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence>], Accessed 14 February 2026
4. Conn, A., *Existential Risk*, Future of Life Institute, Available at:
[<https://futureoflife.org/existential-risk/existential-risk/>], Accessed 9 February 2026
5. Future of Life Institute, Pause Giant AI Experiments: An Open Letter, Available at:
[<https://futureoflife.org/open-letter/pause-giant-ai-experiments/>], Accessed 4 February 2026
6. Hoxtell, W., Global AI Regulation at a Time of Transformation: The Council of Europe's Framework Convention on Artificial Intelligence, EnsuredEurope, Available at:
[<https://www.ensuredEurope.eu/publications/global-ai-regulation>], Accessed 14 February 2026
7. Organization for Economic Co-operation and Development launched, *The OECD.AI Policy Navigator*, Available at:
[<https://oecd.ai/en/dashboards/overview>], Accessed 13 February 2026
8. Vinge, V., *The Coming Technological Singularity: How to Survive in the Post-Human Era*, Available at:
[<https://ntrs.nasa.gov/api/citations/19940022856/downloads/19940022856.pdf>], Accessed 7 February 2026