v. 05, nº 02 - jul/dec 2025
ISSN 2763-8685

LATIN AMERICAN JOURNAL OF EUROPEAN STUDIES

BRIDGE WATCH
JEAN MONNET POLICY DEBATE

LACES
Latin American Center of European Studies

Co-funded by the European Union

TABLE OF CONTENTS

https://dx.doi.org/10.51799/2763-8685v5n2006

BETWEEN INNOVATION AND RISK:

Regulating artificial intelligence under Brazilian Bill no. 2,338/2023 and The Eu Ai Act (Regulation (Eu) 2024/1689 - Challenges for the protection of fundamental rights[1]

Álvaro Sampaio Corrêa Neto[2]
Cristina Mendes Bertoncini Corrêa[3]
Desirré Dornelles de Ávila Bollmann[4]

ABSTRACT: Artificial Intelligence (AI), particularly generative AI, is considered revolutionary and disruptive technology and, as such, has the potential to rapidly replace other technologies. Its rapid installation in social reality represents a challenge to the regulatory capacity of economic blocs and States. While it expands human capabilities in knowledge production, task automation, and content creation on an unprecedented scale, its use raises macro-structural risks related to fundamental rights, democracy, the world of work, and the environment. This article analyzes the European Artificial Intelligence Regulation (EU AI Act 2024/1689) and Brazilian Bill № 2338/2023 and aims to identify mechanisms for addressing the risks associated with AI and evaluate their sufficiency in protecting fundamental rights. Based on Ulrich Beck's concept of a risk society, we examine the difference between predictive (specialized) AI and generative AI and then analyze the risks to fundamental rights posed by these technologies. A significant convergence was found between the European and Brazilian regulatory frameworks, both grounded in risk-based proportional regulation and the centrality of fundamental rights protection. However, relevant gaps were identified: in the European model, challenges persist in regulating general-purpose (GPAI) systems and in the practical implementation of supervisory mechanisms; and in the Brazilian model, the

1. Álvaro Sampaio Corrêa Neto, Cristina Mendes Bertoncini Corrêa, and Desirré Dornelles de Ávila Bollmann, "Between Innovation and Risk: Regulating Artificial Intelligence under Brazilian Bill № 2,338/2023 and the EU AI Act (Regulation (EU) 2024/1689 – Challenges for the Protection of Fundamental Rights)," *Latin American Journal of European Studies* 5, no. 2 (2025): 140 et seq.
2. Master's degree in Production Engineering from the Federal University of Santa Catarina (UFSC) and a bachelor's degree in Computer Science from UFSC. Currently working as Artificial Intelligence Advisor at the Regional Electoral Court of Santa Catarina (TRE/SC). https://orcid.org/0009-0009-2695-9256.
3. PhD in Law (Federal University of Santa Catarina - UFSC). Postdoctoral Fellow in Artificial Intelligence at the University of Brasília (UnB). Professor at the Law School (Under-graduate & Graduate Programmes) - Federal University of Santa Catarina. Researcher in the Dr.IA and LabDr.IA research group (UnB). Lawyer. https://orcid.org/0009-0008-3795-682X.
4. Master's degree in Law from the University of the Itajaí Valley in Santa Catarina, Post Graduation Specialization in International Relations for Mercosur from the University of Southern Santa Catarina,Professor of the lato sensu postgraduate program in Labor Law at the Association of Labor Magistrates of the 12th Region in partnership with the University of Planalto Sul (Uniplac), Member of the Santa Catarina Academy of Labor Law, Labor judge of the regional labor court of Santa Catarina (TRT 12). https://orcid.org/0009-0005-6480-7836.

absence of a clearly defined competent authority, insufficient technical standards for explainability, and deficiencies in the proportionality of sanctions. We concluded that, although consistent at the normative level, the European and Brazilian regimes still require additional regulation, greater institutional capacity for implementation, and stronger practices by independent authorities to ensure the effectiveness and immediate justiciability of fundamental rights in the context of artificial intelligence and its risks.

KEYWORDS: Generative Artificial Intelligence (Generative AI); Fundamental Rights; EU AI Act; Brazilian Bill № 2,338/2023;

ENTRE A INOVAÇÃO E O RISCO: A REGULAÇÃO DA INTELIGÊNCIA ARTIFICIAL SOB A ÓTICA DO PROJETO DE LEI BRASILEIRO Nº 2.338/2023 E DA LEI DE IA DA UE (REGULAMENTO (UE) 2024/1689) - DESAFIOS PARA A PROTEÇÃO DE DIREITOS FUNDAMENTAIS

RESUMO: A Inteligência Artificial - IA e especialmente a IA Gen é considerada uma tecnologia revolucionária e disruptiva e, como tal, tem vocação para substituir outras tecnologias em alta velocidade, representando sua rápida instalação na realidade social um desafio à capacidade regulatória dos blocos econômicos e dos Estados. Ao mesmo tempo em que amplia a capacidade humana de produzir conhecimento, automatizar tarefas e criar novos conteúdos em escala inédita, sua utilização levanta riscos macroestruturais relacionados à direitos fundamentais, à democracia, ao mundo do trabalho e ao meio ambiente. O presente artigo realiza análise entre o Regulamento Europeu de Inteligência Artificial (AI Act – Regulamento (UE) 2024/1689) e o Projeto de Lei brasileiro nº 2.338/2023, buscando identificar mecanismos de enfrentamento dos riscos associados à IA e avaliar sua suficiência na proteção de direitos fundamentais. Partindo da noção de sociedade de risco formulada por Ulrich Beck, examina-se a diferença entre a IA Preditiva (Especializada) e a IA Generativa, para após verificar os riscos aos direitos fundamentais derivados dessas tecnologias. Constatou-se, após, significativa convergência entre os modelos regulatórios europeu e brasileiro, ambos estruturados na regulação proporcional por risco e na centralidade da tutela de direitos fundamentais. Todavia, identificaram-se lacunas relevantes, destacando-se, no caso europeu, desafios quanto à regulação de sistemas de propósito geral e à implementação prática da fiscalização e no caso brasileiro, indefinição da autoridade competente, insuficiência dos padrões técnicos de explicação e na dosimetria sancionatória. Conclui-se que, embora possuam consistência no plano normativo, os regimes europeu e brasileiro dependem de regulamentação adicional, capacidade institucional para operacionalização e práticas de autoridades independentes para assegurar a efetividade e justiciabilidade imediata dos direitos fundamentais frente aos riscos da inteligência artificial.

PALAVRAS-CHAVE: Inteligência Artificial Generativa; Direitos Fundamentais; AI Act; PL 2.338/2023;

INTRODUCTION

Artificial Intelligence - AI, especially generative AI, is considered a revolutionary and disruptive technology and, as such, has the potential to replace other technologies at high speed, representing a challenge to the regulatory capacity of economic blocs and States.

The speed with which artificial intelligence is spreading, combined with its technical opacity, raises an immediate question for the law: are the mechanisms established in emerging legislation sufficient to address the risks of artificial intelligence without stifling its disruptive nature?

This article aims to analyze the mechanisms for addressing the risks of artificial intelligence set out in the European AI Act (Regulation (EU) 2024/1689) and in Brazilian Bill № 2338/2023, evaluating their points of convergence, shortcomings, and the adequacy of normative frameworks to safeguard the effectiveness and immediate justiciability of fundamental rights.

The article is structured in four stages. The first presents the foundations of predictive and generative AI, highlighting their main differences and laying the groundwork for the subsequent discussion of AI-related risks. The second examines the macrostructural risks posed by generative AI, based on Ulrich Beck's theory of risk society, correlating them with generations of fundamental rights. The third analyzes the mechanisms for addressing risks set out in the European AI Act, with emphasis on the regulatory classification, prohibited practices, high-risk obligations, and transparency instruments. The fourth examines the regime proposed by Bill 2338/2023, and addresses the rights of affected individuals, the categories of excessive and high risk, algorithmic impact assessments (AIAs), and the civil liability regime.

Finally, the article concludes by discussing the structural weaknesses of both frameworks by assessing their normative sufficiency and identifying the conditions necessary for the mechanisms set out in these instruments to effectively protect fundamental rights in a context of technological and social uncertainty.

1. ARTIFICIAL INTELLIGENCE AND THE REGULATORY CHALLENGE

According to American researcher Andrew Ng[5] and Chinese developer Kai-Fu Lee,[6] AI represents the new electricity, "the electricity of the 21st century."

With this analogy, the researchers sought to express that Artificial Intelligence plays a role in today's society similar to that played by electricity in the 19th century. Initially, electricity merely replaced candles for lighting, but today it is indispensable. In other words, AI is already integrated into our daily lives, and its presence is expected to increase exponentially.

Currently, AI applications are seen in facial recognition, personal assistants, content recommendation systems, and social networks where content is suggested by AI algorithms.

In this way, just as electricity drove dramatic changes in the past, Artificial Intelligence is driving a transformation of similar magnitude in our time, but at a significantly faster pace.

Artificial Intelligence is considered a revolutionary and disruptive technology.

In the 1990s, J. Bower and C. Christensen identified the process of "disruptive innovation" as a phenomenon in the business sector, in which a smaller, technology-driven company revolutionizes a market segment to the point of upending the business of a larger, well-established competitor, who is then unable to maintain its leadership in the face of radical technological changes in the market.[7] In other words, it describes an innovation that creates a new market and a new value network, eventually displacing established markets, products, and alliances.

5. In 2017, the co-founder of Google Brain and adjunct professor of Computer Science at Stanford University made this statement that became famous among his peers. https://www.gsb.stanford.edu/insights/andrew-ng-why-ai-new-electricity .
6. Kai-Fu Lee, *Artificial Intelligence: how robots are changing the world, the way we love, relate, work and live* (Globo Livros, 2019).
7. Joseph L. Bower and Clayton M. Christensen, "Disruptive technologies: catching the wave" *Harvard Business Review* 73, no. 1 (1995): 43–53, https://www3.yildiz.edu.tr/~naydin/MI2/lectures/Reading/Disruptive%20Technologie%20Catching%20the%20Wave.pdf

The concept of revolutionary technology is one that causes major transformations to reality.[8]

In short, every disruptive technology is, in some way, revolutionary in its ultimate impact, but not every revolutionary technology is disruptive. Artificial Intelligence, specifically, is an example that manages to fit both definitions, which explains why it is considered one of the most important innovations of our time.

Ribeiro[9] emphasizes that, ultimately, innovations that radically reformulate ways of doing things are intrinsic to human and societal existence and evolution, which are becoming more complex and, as a result, depend on and foster inventive creativity capable of optimizing tasks and meeting new challenges in previously unimaginable ways.

The impact of disruptive and revolutionary innovations on social reality is explosive and highlights the difficulty for regulatory policy to examine and contain potential positive and negative aspects linked to the introduction of innovative technology, given the speed at which it is implemented within society.

For Feigelson, "law, in many cases, trails behind facts which, in the context of disruptive dynamics, is even more noticeable given that such models advance within society at a speed that is incompatible with normative changes."[10] In other words, technology moves like drones, while regulations move like steam engines.

Following this line of thought, Leonardo Coelho Ribeiro proposes two reasons why innovations challenge the law: (i) they tend to operate in blind spots, gaining a competitive advantage over established market players to achieve faster and more significant economic gains; and (ii) they expand their activities exponentially and rapidly to become "*too big to ban.*"[11]

8. José Benedito Lázaro da Silva, "The disruptive effect of innovations and technologies on legal and social sciences," in *Regulation and new technologies,* coordinated by Rafael Veras Freitas, Leonardo Coelho Ribeiro and Bruno Feigelson, (Forum, 2018), 159.
9. Leonardo Coelho Ribeiro, "The instrumentality of administrative law and the regulation of new disruptive technologies," in *Regulation and new technologies*, coord. Rafael Veras Freitas, Leonardo Coelho Ribeiro and Bruno Feigelson (Forum, 2018), 72.
10. Bruno Feigelson, "The relationship between disruptive models and the law: establishing a methodological analysis based on three stages," in *Regulation and new technologies*, coord. Rafael Veras Freitas, Leonardo Coelho Ribeiro and Bruno Feigelson (Forum, 2018), 51-52.
11. Ribeiro, "*Instrumentality of law*" 76.

The regulatory challenges posed by technological innovations across multiple sectors now transcend the borders of states and territorial blocs, as globalized societies experience political, technological, economic, and cultural interaction at an unimaginable speed driven by advances in communication technologies.[12]

Giddens observed that, far from the modern ideal of stability, what we see today is a world out of control, with "some of the influences that were previously supposed to make life predictable and safe for us, including the progress of science and technology, often having the opposite effect" [13]- confronting societies with new and unprecedented risks.

In fact, Ulrich Beck defines late modern society as a risk society, in which "problems and distributive conflicts of the scarcity society overlap with the problems and conflicts arising from the production, definition and distribution of scientifically and technologically produced risks."[14]

Disruptive innovations, due to their unique characteristics of transformation to the point of destroying not only economic but also previous social patterns, and the very speed with which they become a reality in people's lives, are part of the risk spectrum to be assessed in the societal model, requiring confrontation through both an effective action plan and monitoring and reflection.

From a regulatory perspective, two main guiding principles are required: clarity regarding which fundamental rights must be protected against potential risks associated with the use of disruptive technologies, and the need to map and adopt preventive measures to mitigate their impact on those fundamental rights.

In parallel, it is necessary to establish specific regulations that are sufficiently flexible and equipped with mechanisms to adapt to a disruptive reality.

However, before we move on to an analysis of regulatory issues, it is crucial to understand the technological nature of AI.

12. For Anthony Giddens, globalization is a complex set of processes, which also involves the partial derogation of the sovereignty of States, in a context of intensification of political, technological, economic, cultural and ecological influences and interactions, which are not necessarily equitable or balanced (Anthony Giddens, *World Out of Control,* (Record, 2003), 17-29.
13. Giddens, "World *Out of Control*" 14.
14. Ulrich Beck, "Risk *Society: Towards a Different Modernity*," (Editora 34 Ltda, 2010). 23.

Unlike traditional software, which operates based on clear, logical rules pre-established by a human programmer (if x, then y), an AI system is "trained."[15] This process involves training a mathematical model—usually a neural network—on a large volume of data, enabling it to adjust its internal weights automatically to recognize patterns and subsequently make predictions. The result is a system that operates probabilistically, not deterministically: it does not "*know*" the correct answer but calculates the most probable one based on the data on which it was previously trained. This transition from deterministic logic to probabilistic inference, using vast amounts of data, poses significant regulatory challenges as it shifts the causal nexus from direct human command to an opaque, large-scale learning process.

In this context of probabilistic inference, it is important to distinguish between two functional categories of AI, each characterized by different uses and training methods. On the one hand, we have predictive—or specialized— AI, also known as single-purpose AI, designed to perform specific tasks, such as classification or prediction. These AIs are typically trained on relatively large datasets but are normally focused and labeled for specific tasks, such as a collection of medical images for disease detection or a database of contracts for identifying risk clauses. Such systems operate on existing data to deliver a specific analysis or result within a predefined scope.

In contrast to this single-purpose approach, generative AI is not trained for a single purpose but rather to create new and original content that mimics the data on which it was trained.[16] Training a generative system involves vast and heterogeneous volumes of data— often in the order of terabytes or petabytes of text, images, and videos. These data are typically extracted from the internet and use self-supervised learning techniques in which the model itself learns

15. David Fernandez Llorca, Vicky Charisi, Ronan Hamon, Ignácio Sanchez and Emilia Gomez, "Liability Regime in the Age of AI: A Use-Case Driven Analysis of the Burden of Proof , *Journal of Artificial Intelligence Research,* no 76 (2023), 613, https://doi.org/10.1r.1.14565.

16. Cristina Mendes Bertoncini Correa, Débora Bonat, Mariane Carolina Gomes da Silva Rocha, Fabricio Ataides Braz, Nilton Correia da Silva, Luciana Nishi, Eduardo Camargo de Siqueira. "The challenges in the development of artificial intelligence applied to law: an analysis of the Osíris project". *Digital Democracy and Electronic Government Journal* 1, no. 24, Florianópolis (2025) 56-69

underlying structures and patterns from the data without the need for explicit human labeling.

The fundamental difference, therefore, lies in the scope of its function, the nature of its output, and the demand for data and training methods: predictive or specialized AI is analytical, trained on specific data and normally labeled for recognition and prediction, whereas generative AI is synthetic, trained on massive volumes of unstructured data to create unprecedented digital artifacts. Therefore, the training and use of generative AI raise and drive new and complex debates on the impact of this disruptive and revolutionary technology on society.

One of the most immediate and legally significant impacts of generative AI lies in its extraordinary speed and ease of use. Whereas the creation of sophisticated digital content once required specialized technical knowledge and considerable time, today, anyone with internet access can generate complex texts, images, audio, or videos in seconds. This transformation has been enabled by intuitive interfaces that operate through prompts—simple commands in natural language—eliminating the need for programming, design, or editing skills. This simplicity of use may explain the speed with which these tools have spread: platforms like OpenAI's ChatGPT, which reached 100 million users in about two months, have democratized access to unprecedented content production capacity. However, this combination of agility and simplicity also magnifies the risk for misuse, making the creation of realistic deepfakes or the orchestration of mass disinformation campaigns a trivial task and extremely challenging to detect.[17]

In this context of synthetic content proliferation, vulnerability to such manipulative use is unevenly distributed across society, with disproportionate exposure to specific groups. Individuals with low AI literacy and who are unfamiliar with the technology's ability to generate content that is true to life become prime targets

17. Momina, Marriam Nawaz Masood, Khalid Mahmood Malik, Ali Javed, Aun Irtaza, and Hafiz Malik, "Deepfakes Generation and Detection: State-of-the-Art, Open Challenges, Countermeasures, and Way Forward." Applied Intelligence 53 (4) 3974 (2022).

as they tend to accept texts, images, and videos as authentic representations of reality, without the skepticism necessary to question their source.[18]

Even more critically, children and adolescents represent a highly vulnerable group. Being in the midst of cognitive and moral development, their capacity for discernment is intrinsically limited. This vulnerability is exploited through processes of early adultification and hyper-sexualization in which they are exposed to adult behaviors, aesthetics, and responsibilities, often with sexual connotations, as evidenced by recent reports on algorithms that actively promote such content.[19]

In addition to the issues already addressed, the large-scale implementation of artificial intelligence creates systemic challenges that deserve in-depth analysis.

2. MAPPING THE MACROSTRUCTURAL RISKS OF GENERATIVE ARTIFICIAL INTELLIGENCE TO FUNDAMENTAL RIGHTS

As already highlighted in the previous section, Ulrich Beck considers late modern society a risk society.[20] The risk Beck refers to is systematically defined, often invisible, but capable of being revealed through knowledge. It can be altered, diminished, or increased and is, therefore, open to social processes of definition.[21]

Yet, in essence, it refers to the future extent of currently foreseeable damage, to threats projected into the future.[22]

From the moment generative AI is capable of creating realities and even virtual people in a hyper-realistic way that is indistinguishable from reality itself, to interact and talk to human beings, and even to simulate emotions within these

18. Gergely Ferenc Lendvai and Gergely Gosztonyi, "Deepfake and Misinformation – What Can Doing the Right in Front of the Fake News Created By Deepfake?" IDP Derecho Internet Magazine y Politicos, no. 41 (2024). https://doi.org/10.7238/idp.v0i41.427515.
19. https://www.cnnbrasil.com.br/tecnologia/adultizacao-o-que-meta-e-google-estao-fazendo-apos-denuncias-de-felca/#google_vignette
20. Beck, "*Risk Society*" 23.
21. Beck, "*Risk Society*" 27.
22. Beck, "Risk *Society*" 39.

interactions while learning from them and reproducing content and realities, it becomes evident that this technology poses a threat to fundamental rights.

Indeed, on the subject of human and fundamental rights, JJ Gomes Canotilho explains:

> The terms "*human rights*" and "*fundamental rights*" are often used synonymously. However, according to their origin and meaning, they can be distinguished as follows: *human rights* are rights that apply to all peoples and at all times (natural law-universalist dimension); *fundamental rights* are human rights that are legally and institutionally guaranteed and limited in space and time. *Human rights* derive from human nature itself and therefore have an inviolable, intertemporal, and universal character, whereas *fundamental rights* are those objectively valid in a concrete legal order.[23]

For this author, fundamental rights fulfill the functions of: a) defense or freedom; b) social provision; c) protection against third parties; and d) non-discrimination.[24]

With regard to the function of defense rights of citizens, fundamental rights, on the one hand, "constitute, on an objective legal level, norms of negative competence for public authorities, fundamentally prohibiting interference in the individual legal sphere"[25] and, on the other hand, "imply on a subjective legal level, the power to positively exercise fundamental rights (positive freedom) and to demand omissions from public authorities to avoid harmful infringements by them (negative freedom)."[26]

Rights to benefits are those held by individuals to obtain assistance through the State.[27]

The function of protection against third parties consists of the "duty of the State to adopt positive measures aimed at safeguarding the exercise of fundamental rights against disruptive or harmful activities carried out by third parties."[28]

23. José Joaquim Gomes Canotilho, *Constitutional Law and Constitutional Theory* (Almedina, 2000), 387.
24. Canotilho, "*Constitutional Law*" 401.
25. Canotilho, "*Constitutional Law*" 401.
26. Canotilho, "Constitutional *Law"* 401.
27. Canotilho, "Constitutional *Law"* 402.
28. Canotilho, "Constitutional *Law"* 402.

The right to non-discrimination is the function of fundamental rights to ensure that the State treats its citizens as fundamentally equal citizens.[29]

George Marmelstein, on the other hand, attributes the theory of the "*generations - evolution - of fundamental rights*" to the Czech jurist Karel Vasak, which, based on the ideals of the French Revolution, are distributed as follows: a) the first generation of rights comprises civil and political rights founded on freedom, which emerged from the bourgeois revolutions; b) the second generation consists of economic, social and cultural rights, founded on equality and driven by the Industrial Revolution and the social problems it caused; c) finally, the last generation encompasses solidarity rights, especially the right to development, peace and the environment, crowning the triad with fraternity. This notion gained strength after the Second World War and particularly after the Universal Declaration of Human Rights in 1948.[30]

Bobbio rightly states that the Modern Age corresponds to the Age of Rights– understood as the era that elevates the former servant to the status of citizen, for whom democracy is an essential condition, [31] and one which recognizes the structuring of human rights into generations, even pointing to the existence of a fourth generation of rights related to biotechnology and bioethics.[32]

The authors of the article are aware that, today, many authors prefer to replace the expression "generations of fundamental rights" with "dimensions of fundamental rights". For this reason, both expressions will be used throughout the article.

From this brief overview of fundamental rights and their respective generations and dimensions, it is possible to identify potential macrostructural risks of generative AI to these rights.

Saetra[33] analyzes four potential macrostructural advantages and risks associated with the widespread use of this AI model.

29. Canotilho, "Constitutional *Law"* 402.
30. George Marmelstein, *Fundamental Rights Course*, (Altas, 2008), 42.
31. Norberto Bobbio, *The Age of Rights* (Campus, 1992), 65.
32. Bobbio, "Age *of Rights*" 6.
33. Saetra, "Generative *AI*".

The first refers to the impacts on democracy and political stability.

For the author, while there may be advantages in adopting generative AI for democratic management, it is also known that this technology can generate unlimited political content for dissemination across digital networks, with an increased risk of "fake news".

According to Saetra:[34]

> Fake news is a concern, and this can be both texts and generated videos in which real people or situations are presented in new and imaginary ways (deepfakes). When AI floods the information sphere with new content, there are, for example, fears that people will lose track of what is true and what is not, or that polarization will increase. [35] (authors' translation).

In this regard, Nik Hynek, Beata Guavurova and Matus Kubak[36] state that the emergence of *deepfakes* has sparked significant legal debate concerning regulation and challenges brought about by AI, with profound legal implications as it can spread disinformation and influence elections and deeply impact democracy and the rule of law.

DeepFake, which involves simulating emotions in artificial interactions, poses risks to fundamental rights, such as freedom, privacy, equality, and non-discrimination.

Democracy, as a condition for the existence of human rights, can be destabilized by the use of AI to fuel social polarization, *fake news* and *deepfakes,* threatening the regularity of electoral processes, the reliability of institutions, and the very social perception of truth.

A second impact relates to transformations in the world of work, notably the replacement of workers by generative AI. While it is accepted that the introduction of generative artificial intelligence can simplify processes, reduce the

34. Saetra, "Generative *AI*".

35. Fake news is one concern, and this could be both text and generated videos where real people or situations are presented in new and imagined ways (deepfakes). When AI floods the information sphere with new content, there are, for example, fears that people will lose their grasp of what is true or not, or that we'll experience increased polarization.

36. Nik Hynek, Beata Guavurova , and Matus Kubak, "Risks and benefits of artificial intelligence deepfakes: Systematic review and comparison of public attitudes in seven European Countries," *Journal of Innovation & Knowledge* 10, no. 5, (2025), https://www.sciencedirect.com/science/article/pii/S2444569X25001271 .

time required for various services, and provide *insights* into work, its impact on unemployment must be assessed in light of the adoption in 2015 by the Member States of the United Nations of the 2030 Agenda for Sustainable Development and the commitment to Sustainable Development Goal № 8, which addresses decent work and economic growth[37.38]

The disruptive nature of this technology raises concerns about structural unemployment and the precarization of labor relations, as the introduction of generative AI into the labor market has a direct effect on second-generation fundamental rights, whose foundation lies precisely in the promotion of substantive equality and the guarantee of minimum social conditions for a dignified existence.

A third impact concerns how generative AIs are trained, that is, how they are fed and the nature of the data used. Whether historical or synthetic—the latter often derived from the former—their use can result in maintaining the *status quo* and stifling the possibility of social change on the one hand, and in the risk of reproducing and even amplifying discrimination and prejudice on the other, often in subtle ways and within new and opaque contexts.[39]

At the legal level, these risks are potentially serious as they directly impact the logic of fundamental rights in various dimensions.

For this reason, Fabiano Hartmann and Debora Bonat argue that the logic of mere substitution in the technological expansion of AI over law is inadequate, since such expansion demands a prior commitment to respect for and commitment to fundamental rights.[40.]

It is worth noting that algorithmic opacity represents a true Pandora's box. It undermines social oversight, the attribution of responsibility (accountability), and reparation, all within a potentially harmful system.

37. Saetra, "*Generative AI*".
38. UN - United Nations Brazil, *Sustainable Development Goals: Decent work and economic growth*, (2025), https://brasil.un.org/pt-br/sdgs/8 .
39. Saetra , "*Generative AI*".
40. Debora Bonat and Fabiano Hartmann Peixoto, " GPTs and Law: probable impacts of generative AIs on Brazilian legal activities"**,** *Sequence: Legal and political studies* 44, (2023) 4-5.

The risk, therefore, lies not only in the explicit reproduction of prejudices but also in subtle forms inherent to the system.

In this context, instead of contributing to the realization of the fundamental right to equality in its material dimension, generative AI threatens to become yet another instrument for perpetuating historical inequalities.

It is important to emphasize that fundamental rights play a crucial protective role, requiring not only abstention on the part of the State but also active protection of citizens, extending both vertically and horizontally, so that action must also be taken to prevent the compromise of these rights by private actors.

Finally, the environmental impact of generative AI cannot be ignored.

It should not be overlooked that, as a disruptive technology, a world without access to generative artificial intelligence is no longer conceivable. However, the processes involved, whether for training or supply of this type of technology, consume a highly significant amount of energy drawn from the Earth's environmental resources, thereby generating a significant carbon footprint.[41]

Zewe, in an article published in MIT News, stated that:

> Scientists estimate that data center power demand in North America increased from 2,688 megawatts at the end of 2022 to 5,341 megawatts at the end of 2023, driven in part by the demands of generative AI. Globally, data center electricity consumption increased to 460 terawatt -hours in 2022. This would make data centers the 11th-largest electricity consumer in the world, between Saudi Arabia (371 terawatt -hours) and France (463 terawatt -hours), according to the Organization for Economic Cooperation and Development (OECD). By 2026, data center electricity consumption is expected to approach 1,050 terawatt -hours (which would place data centers fifth on the global list, between Japan and Russia).[42] authors' translation.

41. Saetra, "Generative *AI*".

42. Adam Zewe, "Explained: The Environmental Impact of Generative AI," *MIT News on Campus and around the world* (2025), https://news.mit.edu/2025/explained-generative-ai-environmental-impact-0117. Scientists estimate that energy demand from data centers in North America will increase from 2,688 megawatts at the end of 2022 to 5,341 megawatts at the end of 2023, partly driven by the demands of generative AI. Overall, electricity consumption by data centers will increase to 460 terawatt -hours in 2022. This would make data centers the 11th largest consumer of electricity in the world, between Saudi Arabia (371 terawatt -hours) and France (463 terawatt -hours), according to the Organization for Economic Cooperation and Development (OECD). By 2026, data center electricity consumption is expected to approach 1,050 terawatt -hours (which would place data centers fifth on the global list, between Japan and Russia).

The author further highlights that the sector is currently on an environmentally unsustainable path and that, given the speed of technological advances, scientists have not yet been able to measure or fully comprehend the environmental compensations required.

When confronted with the notion of third-generation fundamental rights— among which one of the most important is the right to an ecologically balanced environment— these data directly threaten the effectiveness of this right, both individually and on a collective and intergenerational level.

In short - and as Yudkowsky points out - "by far, the greatest danger of artificial intelligence is that people prematurely conclude that they understand it"[43] (authors' translation). This leads to an underestimation of the risks compared to the benefits of generative AI, which stem largely from the tendency to anthropomorphize this technology and to presume that it will necessarily be a Friendly Artificial Intelligence.[44]

The contours of generative AI production together with its impacts and risks strongly indicate the need for a regulatory framework that mitigates and manages the risks that directly impact fundamental rights.

3. RISKS TO FUNDAMENTAL RIGHTS AND MITIGATION MECHANISMS LISTED IN EU REGULATION 2024/1689 – (THE AI ACT)

Regulation (EU) 2024/1689 - the AI Act - entered into force in the European Union on August 1, 2024. According to item (1) of the Explanatory Memorandum, the Regulation aims to improve the functioning of the internal market by standardizing the rules that govern artificial intelligence systems (AI systems) within the European Union and promoting the adoption of Artificial Intelligence (AI).[45]

43. By far the greatest danger of Artificial Intelligence is that people conclude too early that they understand it.

44. Eliezer Yudkowsky, *Artificial Intelligence as a Positive and Negative Factor*, (Oxford University Press, 2008) 308–345, https://intelligence.org/files/AIPosNegFactor.pdf.

45. European Union, "Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024," *Journal of the European Union* , (2024), https://eur-lex-europa-eu.translate.goog/legal-content/PT/TXT/?uri=CELEX:32024R1689&_x_tr_sl=en&_x_tr_tl=pt&_x_tr_hl=pt&_x_tr_pto=tc.

The regulation adopts a genuine risk-based scale, which ranges from prohibited practices to minimal risk, consistently anchored in the notion of fundamental rights, especially those outlined in the Charter of the European Union and in international instruments.

Article 3(1) of the Regulation defines an AI system as a machine-based system designed to operate with varying levels of autonomy and that, based on the input data it receives, infers how to generate outputs such as predictions, content, recommendations or decisions capable of influencing physical or virtual environments—without distinguishing between AI and generative AI.

In Article 3(2) of the Regulation, risk refers to the likelihood of harm and the severity of such harm. The following classification can be derived from the Explanatory Memorandum and the articles:

a) use and practices that must be prohibited, where the application of AI is considered abusive and must be absolutely prohibited as it violates fundamental rights;

b) high-risk use and practices, when the use of AI, although not automatically illegal, can produce significant negative impacts on health, safety, workers' rights, access to essential services, and education;

c) use and practices of limited, reduced or low risk, when the impact of the application of AI is restricted, considered as such because it does not substantially affect decision-making or the individual's autonomy.

Some uses of AI must be completely prohibited when they violate fundamental rights and are incompatible with the Democratic Rule of Law, as stated in Recital (28), including:

a) manipulation—including subliminal manipulation—of individuals, and the exploitation of human vulnerability, namely, systems that induce harmful behavior through imperceptible stimuli or sophisticated persuasion techniques, thus compromising autonomy and freedom of choice, as well as exploiting the vulnerabilities of minors, people with disabilities or in situations of socioeconomic

fragility - Recital (29) of the Explanatory Memorandum and in Article 5(a) and (b) of the Regulation;

b) AI programs that perform biometric categorization by using data such as facial features, fingerprints, or other biological characteristics to infer religious beliefs, political opinions, sexual orientation, or aspects of private life - Recital (30) of the Explanatory Memorandum and article 5(g) of the Regulation;

c) the use of AI for social scoring, to evaluate individuals based on data points to establish social hierarchies, with discriminatory outcomes and cause harm, particularly to vulnerable groups, or to assess the risk of committing a criminal offense - Recitals (30) and (31) of the Explanatory Memorandum and Article 5(c) and (d) of the Regulation;

d) use of AI for facial recognition on a large scale, especially in real time and in public spaces, with the risk of permanent surveillance and invasion of privacy, notably for private purposes - Recitals (32) and (43) of the Explanatory Memorandum and Article 5(h) of the Regulation;

e) use of AI to detect emotions, notably in work and education contexts. AI systems aimed at identifying or inferring emotions still lack a reliable scientific basis and, in a context of power asymmetry, may serve to discriminate - Recital (44) of the Explanatory Memorandum and Article 5(f) of the Regulation.

The exceptions provided for in this type of system concern health, criminal prosecution or victim protection, and security in certain cases, subject to judicial authorization or approval by an independent authority, as well as other law enforcement purposes to be regulated by the Member States.

In turn, among the domains identified as high risk in the Explanatory Memorandum Recitals (56) to (61) are those contained in Annex III, and generally refer to:

a) In the area of education, systems that determine access to institutions, monitor students or evaluate performance in an automated manner;

b) at work, systems for recruiting, promoting, evaluating or monitoring workers, with an impact on the livelihood and dignity of the employee;

c) access to essential services, health systems, social security, housing, and public assistance systems, where AI can decide whether to grant or deny vital benefits

d) in the area of justice and security, where the use of AI in judicial proceedings or for law enforcement purposes may affect the fundamental right to the presumption of innocence, the right to action and defense, to a fair trial, as well as individual freedoms;

e) in the management of migration, asylum and border control, as such uses may affect vulnerable groups whose fate depends on the actions of public authorities;

f) in democratic processes, where AI applications can have a potentially significant impact on democracy, the rule of law and individual freedoms.

Finally, Recital (53) refers to limited, reduced, or low-intensity risks, defined as such because they do not substantially affect decision-making or individual autonomy. This includes cases where AI performs merely preparatory tasks, such as classifying texts or documents or detecting past patterns, as well as those supplementary to human activities that have already been completed.

Regarding the risk-addressing mechanisms present in the Regulation for systems whose use directly clashes with fundamental rights, the European Union initially sets out in Article 5 a categorical prohibition that such applications cannot be placed on the market or put into service in the European Union. Reinforcing this prohibition is the provision that market surveillance authorities must monitor compliance and may apply penalties (Articles 63-64); limitations on exceptions, such as biometric identification for security purposes, are subject to judicial/administrative authorization, with an impact assessment and temporal/spatial limits (Recital 35); Article 5, paragraph 1, letter d)) and the integration of the Regulation with other EU regulatory instruments, such as Directive (EU) 2016/680. This strengthens safeguards put in place for biometric data (Recital 38).

With regard to permitted AI applications that are considered high risk, the protection system provides for multi-layer protection.

Article 9 establishes the adoption of risk management, which is mandatory throughout the AI life cycle with periodic reviews, and the inclusion of risks since the AI´s conception, elimination, or hierarchical mitigation (Art. 9, no. 5), and consideration of children and vulnerable groups (Art. 9, no. 9).

To avoid discriminatory biases, Article 10 provides for data governance and requires quality, representativeness, and relevance of training and test data.

Article 11 and Annex IV contain the technical documentation required for compliance of high-risk AI systems. These documents must include information such as purpose, interactions, development methods, *datasets*, performance, and usage instructions for auditing and monitoring purposes. These documents are subject to record-keeping obligations (Articles 12 and 19), requiring automatic *logs* to ensure traceability, and must be retained for a minimum period of six months.

Article 13 details transparency and information obligations, and requires suppliers to provide instructions for proper use, system limits, maintenance conditions, and correct interpretation of results, including guidance to prevent "automation bias" in points of human decision-making

Article 14 already lays down the essential requirement of human oversight (n. 1), such that the AI system must be designed to allow effective human intervention to prevent or minimize risks when used in accordance with its intended purpose or reasonably foreseeable misuse (Art. 14, paragraph 2). This must include procedures for understanding or limiting the system, detecting and correcting anomalies, the ability to decide not to use, override or reverse outcomes, and the activation of a stop button. Certain contexts, such as law enforcement (Annex III, point 1(a), require dual human verification (Art. 14, paragraph 5).

Article 15 establishes that AI systems must be robust against errors, faults and attacks (cybersecurity) throughout their lifecycle and accuracy results must be included in the instructions (Art. 15, paragraphs 1 and 3).

Article 43 sets out the conformity assessment procedures for high-risk systems, namely internal control (Annex VI) or quality management and technical

documentation assessment with a notified body (Annex VII), depending on the existence or application of harmonized standards (Article 40) and/or common specifications (Article 41). Furthermore, Article 49 stipulates that high-risk systems must be registered in a European Union database, and public access to this repository must contain essential non-confidential information for transparency and *accountability* purposes.

The Regulation also provides for post-market monitoring (Art. 72) and incident notification (Art. 73) whereby suppliers are required to monitor the development of AI systems after they have entered the market and to report serious incidents.

Articles 58 through 60 provide for testing environments with appropriate safeguards for fundamental rights, health, and safety in cases where real-world supervised testing is conducted (Articles 58, 59, and 60). Additional data processing is set out in the *sandbox* for public interest under strict conditions (Art. 59).

Article 6, paragraph 3 provides an exception for systems listed in Annex III that do not pose a significant risk to health, safety or fundamental rights, nor significantly influence the outcome of the decision, in which case they may cease to be classified as high risk.

Finally, with regard to reduced, limited or low risk systems, Chapter IV of the Regulation imposes horizontal transparency obligations, namely: a) when systems interact directly with individuals, this condition must be disclosed (unless it would be obvious to a "reasonable user" and subject to law enforcement exceptions) (Art. 50, paragraph 1); b) providers must mark that the content has been artificially generated or manipulated (Art. 50, paragraph 2); c) emotion recognition or categorization systems, where lawful, must inform exposed individuals and process personal data in accordance with the relevant European Union regulations, in particular GDPR/Reg. 2018/1725/Dir. 2016/680 (with police exceptions) (Art. 50, paragraph 3); and d) suppliers who implement systems that generate/manipulate images, audio or video with realistic similarity (deep fakes) must clearly point out this artificiality, except under responsible human

editorial control (there are exceptions for law enforcement and artistic/satirical works) (Art. 50, paragraph 4).

As can be seen, the principle of proportionality applies in this case, such that low-risk systems, although not subject to the same regulatory level of control as high-risk ones, still remain bound by minimum transparency obligations, which allow for a certain degree of user oversight.

Sabrina Kutscher[46] reports that, beginning with the General Data Protection Regulation (Regulation (EU) 2016/679), the European Union initiated a strong regulatory strategy to govern digital technologies, combining the logic of the internal market with the protection of fundamental rights such as privacy and data protection. This was also a political move aimed at strengthening the EU's geopolitical position vis-à-vis *big tech* and consolidating its image as a defender of the free market and citizens' rights. This logic inspired new regulations—such as the Data Governance Act, the Cybersecurity Act, the Digital Services Regulation, the Digital Markets Regulation, and, finally, the AI Act.

According to the author, regulatory resources in the AI Act are fragmented and distributed among different actors—supply chains, certifications, standardization bodies, consumer protection, and fundamental rights entities— as well as existing regulators. She notes that the AI Act combines formal legal rules with multiple *soft law* instruments (codes of conduct, recommendations, guidelines), which makes the regulatory framework more complex.[47]

Martens[48] argues that the AI Act was not drafted with generative AI in mind but rather for single-purpose AI models. For generative AI, however, it is difficult to effectively anticipate the full range of risks, thereby revealing the insufficiency of the current regulatory model.

46. Sabrina Kutscher, "The EU AI Act: Law of Unintended Consequences?" *Technology and Regulation* 2025 (2025): 355-385, https://techreg.org/issue/view/1700.
47. Kutscher "EU *AI Act* ".
48. Bruegel Martens, "The European Union Ai Act: Premature or Precocious Regulation?" *Bruegel* (2024), https://www.bruegel.org/analysis/european-union-ai-act-premature-or-precocious-regulation .

The author also points out shortcomings in regulation, citing as examples: the attribution of responsibility of transparency to models considered low risk only, as well as the existence of exceptions that translate into regulatory loopholes regarding high-risk systems,[49] the classification of systemic risk to high-power models while ignoring potential risks of smaller models; unclear attribution of responsibilities in the usage chains; and a very generic definition of the values to be protected.[50]

Martens, in short, believes that the AI Act, as it stands, does not adequately address the open and systemic nature of general-purpose AI and leaves gray areas of responsibility, imposes high costs with low predictability (especially on small players), and relies on future standards and institutional capacity that do not yet exist. These factors may reduce competitiveness and fail to ensure safety and fundamental rights in the short term.[51]

The AI Act undoubtedly represents a significant regulatory milestone by proposing a risk-based approach to regulate the use of artificial intelligence in the European space. The set of mechanisms provided for in the central articles (Arts. 5, 9 - 15, 43, 49, 50, 72-73) and in the explanatory memorandum recitals (28) to (61) comprise a multilayered system of protection that seeks to balance technological innovation with the preservation of dignity, equality, non-discrimination, privacy, and other fundamental rights enshrined in the Charter of the European Union.

From an efficiency perspective, the Regulation offers valid responses to some of the risks that can be identified in the contexts presented—such as education, work, justice, and access to essential services—by imposing risk management throughout the entire lifecycle, data governance, auditable technical documentation, logs, human oversight, robustness and cybersecurity requirements,

49. Martens "European *Union AI*".

50. Martens "European *Union AI*".

51. Martens "European *Union AI*". For the author "The EU AI Act as it stands is just the start of a long regulatory process. It delegates responsibility to the Commission and its newly AI Office to draft implementation acts and guidelines to address these challenges. These will drive enforcement of the Act and determine to what extent it will be a precocious instrument to stimulate trustworthy AI innovation or a premature innovation-smothering regulation".

public registration of high-risk systems, and post-market monitoring. These mechanisms are capable of mitigating and monitoring negative impacts by establishing effective safeguards against the abusive use of technology.

However, as pointed out by the authors Martens and Kutscher above, this regulation has significant limitations, including the fact that it primarily focuses on specialized AIs trained for a single purpose, rather than specifically for generative AI.

There are also regulatory gaps concerning the absence of sufficiently clear criteria on which human values should guide AI alignment, as well as a reliance on future delegated and implementing acts, which defer the completion of the regulatory framework and make its effectiveness contingent upon the institutional capacity of the new AI Office of the European Commission.

Finally, the Regulation lacks a specific guideline on the environmental impact resulting from the use of generative AI and on how such risks should be mitigated.

4. RISKS TO FUNDAMENTAL RIGHTS AND MITIGATION MECHANISMS LISTED IN BRAZILIAN BILL NO. 2,338/2023

On December 5, 2024, the Brazilian Federal Senate's Internal Temporary Committee on Artificial Intelligence (CTIA) approved the report of the substitute for Bill № 2,338/2023[52], which aims to regulate the use of Artificial Intelligence systems in Brazil, and has since been forwarded to the Chamber of Deputies for deliberation and voting.

As can be seen from the memorandum, the Bill has two main objectives: protecting individuals from the daily impacts of AI—ranging from content recommendations and online advertising to credit and public policy eligibility analyses—and creating legal certainty for innovation and technological development by establishing governance mechanisms and an institutional framework for monitoring and supervision.[53]

52. Brazil. Federal Senate, *Bill № 2338/2023.* Provides for the use of Artificial Intelligence, (Federal Senate, 2024), https://www25.senado.leg.br/web/atividade/materias/-/materia/157233 .

53. Brazil "Bill".

In this regard, Article 2 establishes the guiding principles for the implementation and use of artificial intelligence systems: the centrality of the human person; respect for human rights and democratic values; the free development of personality; environmental protection and sustainable development; equality, non-discrimination, plurality and respect for labor rights; and the combination of technological development with free enterprise. This demonstrates that, although the provision does not expressly refer to risk, fundamental rights are the parameters on which AI systems in Brazil must be created, developed, evaluated, and disseminated.

For this purpose, Article 3 sets out seventeen principles to be observed, beginning with the promotion of inclusive growth, sustainable development, and human well-being, while safeguarding self-determination and freedom of choice. The same article establishes the mandatory and effective participation and supervision of human beings in the creation, development, and dissemination of AI, ensuring non-discrimination, justice, equity, and inclusion. It further emphasizes the duties of transparency, explainability, and auditability, and provides for compliance with due process, including the possibility of challenge and respect for the right to be heard by interested parties, as well as the traceability of decisions to attribute liability to natural or legal persons. Completing this set of principles are the requirements of full reparation for damages, the adoption of preventive and risk mitigation measures, and adherence to the principles of non-maleficence and proportionality between the methods employed and the legitimate purposes of AI systems.

Indicators such as transparency, the right to adversarial proceedings, auditability and traceability, in addition to compensation for damages, demonstrate that the regulation is effectively designed to protect individuals who may suffer harm and establishes mechanisms for individual action against those responsible for the dissemination of AI in the market.

Article 5 enshrines the rights of individuals in relation to the use of artificial intelligence systems, including the right to be informed in advance about in-

teractions, to receive explanations of decisions, to challenge them, to require human participation in the process, and to ensure non-discrimination and the protection of personal data. This paragraph stipulates the duty to clearly inform individuals how these rights may be exercised by AI agents.

Article 6 clarifies that rights related to AI can be invoked both in administrative and judicial proceedings, individually or collectively, under the remedies under Brazilian law.

Articles 7 and 8 address individuals' right to information, including before entering into a contract, and highlight the need for specific information for systems that affect vulnerable groups, as well as transparency in cases of emotion recognition and biometric categorization. They outline the right of affected individuals to request detailed explanations of decisions or predictions, including the criteria used and data considered, with the right to appeal — all free of charge.

Articles 9 to 11 provide for the individual´s to right to challenge AI decisions that significantly impact their interests or rights, including requests for data correction, deletion of abusive information, or review of discriminatory, unreasonable, or inaccurate decisions. Intervention by the affected party is also provided for when decisions have significant legal repercussions. It should be emphasized that, in situations involving risk to life or physical integrity, meaningful human involvement and a final human decision are required.

Article 12 prohibits the use of AI that results in unlawful or abusive discrimination, even indirectly, whether arising from sensitive data or from disproportionate impacts linked to personal characteristics such as race, gender, age, socioeconomic status, or religion.

Concerning risk mapping, the Bill classifies the risks arising from the application of AI into two broad groups: excessive risk and high risk. No express mention of reduced-risk, limited-risk, or low-risk systems is included. These systems are subject to the general rules of the Bill, which primarily establish transparency, governance, and data protection requirements, and adopt techniques and standards that prioritize privacy and information security.

Excessive risk corresponds to the AI practices deemed incompatible with the legal order and human dignity and must therefore be absolutely prohibited.

Article 12 defines that AI systems are prohibited if they:

> a) use subliminal techniques capable of distorting human behavior in a manner that is detrimental to health, safety or fundamental rights (Art. 12, I).
>
> b) exploit the vulnerabilities of specific groups, such as children, the elderly and people with disabilities (Art. 12, II).
>
> c) are employed by public authorities to carry out universal social classification of individuals (Art. 12, III).
>
> d) implement real-time biometric recognition in public spaces, except in restricted cases, provided for in federal law, with judicial authorization and for cases of individualized criminal prosecution, serious crimes or search for victims (Art. 12, IV, §§ 1º-3º).

High risk encompasses AI systems that, while not prohibited, pose significant potential risks to health, safety, and fundamental rights and therefore require control mechanisms. Article 13 of the Bill, supplemented by Annex I, identifies as high risk, among others, systems applied in: education and vocational training, for determining access, monitoring, or evaluating students (Article 13, I); labor and employment, such as in recruitment, promotion, surveillance, and employee evaluation processes (Article 13, II); access to essential services, such as health, housing, and social assistance (Article 13, III); credit assessment and debt capacity (Article 13, IV); public safety and criminal justice, including recidivism predictions, criminal investigations, and analyses (Article 13, V); migration and border management, with an impact on vulnerable groups (Art. 13, VI); the operation of autonomous vehicles and critical infrastructure (Art. 13, VII-VIII) and the health field, with assisted diagnoses or automated interventions (Art. 13, IX).

In terms of risk coping mechanisms, the Bill opts for a general prohibition regarding excessive risk systems (Articles 14 to 16).

Within the scope of high-risk systems (Arts. 17 to 21), the Bill establishes a governance framework that requires full documentation of the life cycle, maintenance of records (logs), accuracy and robustness tests, bias management, explainability and significant human oversight.

In the public sector, particular emphasis is placed on the requirements for prior consultations and hearings, access protocols and the transparency of preliminary assessments.

There is also provision for carrying out AIA (Art. 22 - 26), based on the precautionary principle, which constitutes an innovative measure.

The efficiency of this prediction, however, depends primarily on more detailed and operational regulations, especially with regard to generative AI and general-purpose models, which do not have specific regulations in the Brazilian framework.

As previously stated, the Bill contains no express reference to reduced-, limited-, or low-risk systems; the general provisions of the Bill therefore apply to such cases.

If we return our focus to the people, the Bill lays out strict liability for high and excessive risk systems with the presumption of fault in other cases (Arts. 27 to 29). This ensures full compensation for victims harmed by AI.

The Bill also provides for the mandatory reporting of serious incidents (Art. 31) and sets graduated administrative sanctions. These include warnings, fines of up to R$50 million or 2% of national revenue, suspension and cessation of activities (Arts. 36 and 37).

It also provides for the creation of a public database with information on high-risk systems and strengthens the mechanisms for holding those involved *accountable*.

A comparison between the mechanisms set out in the EU AI Act and Bill № 2,338/83 shows that, with respect to artificial intelligence systems deemed unacceptable or of excessive risk, there is strong regulatory intervention, as both frameworks expressly prohibit the development and deployment of systems of this nature.

In the case of Bill № 2,338/23, the risk categorization of AI systems is conducted through a preliminary assessment, as defined in Article 13. Article 17 provides a list of AI systems considered high-risk, based on their purposes. To classify

a system as high-risk, the Bill uses two approaches: (i) if one of the system's purposes is listed in Article 17; and (ii) through an analysis based on quantitative and qualitative criteria set out in Article 18, which manages the updating of the high-risk AI systems list by the competent authority.

The hypotheses of high-risk systems in the European Regulation are set out in Article 6, supplemented by Annex III, which identifies two main categories of high-risk AI systems: (i) AI systems intended for use as safety components of products subject to third-party conformity assessment; (ii) other autonomous AI systems with implications primarily for fundamental rights, explicitly listed by their area of application in Annex III. Similar to Bill № 2,338/23, the European framework also allows for the list of high-risk AI use cases to be updated as it contains a limited number of AI systems whose risks have already materialized or are likely to materialize in the near future.

FINAL CONSIDERATIONS

Artificial intelligence is a revolutionary and disruptive technology, evolving from the theoretical foundations of thinking machines to neural networks and, more recently, to the leap of large-scale generative and language models.

This context reveals, on the one hand, the speed and breadth of AI's socio-economic impacts and, on the other, the inherent gap between innovation and regulatory response.

Based on the social theory of risk, it follows that the diffusion of AI - and generative AI in particular - produces externalities that affect fundamental rights across different "generations" (or dimensions), including democracy and public deliberation, non-discrimination and material equality, data protection and privacy, decent work and an ecologically balanced environment. This context calls for risk-based regulation that is able to combine prevention, mitigation, accountability and transparency throughout the life cycle of systems.

Building on these premises and focusing on the societal risks of AI, the European AI Act presents a regulatory response that ranges from categorical

prohibitions—practices deemed incompatible with fundamental rights— to a high-risk regime requiring continuous risk management, data governance, auditable technical documentation, logging, and effective human oversight, to horizontal transparency obligations for limited-risk uses including labeling of artificial content and provision of information to the user in human-machine interactions.

Brazilian Bill № 2,338/2023 proposes an equally risk-based model but one that focuses more specifically on the rights of affected individuals—information, explanation, challenge, and human review; protection against discrimination and bias—defining excessive (prohibited) and high risks (subject to reinforced controls) and establishing a set of governance and information security obligations applicable to other cases.

The text outlines algorithmic impact assessment (AIA) as a precautionary tool for sensitive domains, civil liability differentiated by degree of risk, the sanctioning regime and the creation of a public database for high-risk systems, in addition to regulatory sandboxes and rules for text and data mining designed to reconcile protection and innovation.

When the two laws are compared, substantive points of convergence emerge between the regulatory frameworks. Both regimes adopt a risk-based regulatory logic, segmenting prohibited practices (when incompatible with fundamental rights) and establishing a set of reinforced obligations for high-risk applications with emphasis on data governance, minimal explainability, human oversight, robustness and cybersecurity, technical documentation, and traceability. In both systems, the dynamic updating of high-risk lists recognizes that technology evolves more rapidly than legislation and requires mechanisms for ongoing adaptations.

Both the AI Act and the Bill focus on transparency through databases and public information obligations and reaffirm the centrality of fundamental rights as parameters for regulatory design.

An analysis of the frameworks reveals a convergence to establish a cross-cutting regulatory approach that conditions innovation in the protection of individuals and democratic institutions.

However, there are significant distinctions between the models.

The European AI Act is primarily market-oriented and relies on conformity assessment and post-market analysis, combining technical standardization, notification of agencies, and standardized procedures. Alongside more extensive bureaucratic resources for risk assessment, this framework ensures greater control, predictability, and auditability in the European space—albeit at a higher cost, both in terms of government spending and compliance burdens.

In contrast, the Brazilian Bill centers its regulatory approach on the individual, establishing the immediate justiciability of rights and the accountability of actors. It grants individuals explicit rights to request explanations, challenge automated decisions, and demand human intervention—thus reinforcing individual protection within the public sector in particular and essential services.

Conversely, the Bill still lacks an explicit definition of the supervisory authority and minimum technical standards to explain, audit and record systems in a machine-readable manner, which constrain and weaken transparency in practice.

Regarding sensitive domains and uses, the two instruments converge by recognizing high risk in education, work, justice, migration, credit, health, and critical infrastructure, in addition to imposing severe restrictions on biometric recognition in public spaces and prohibiting practices such as "social scoring " and subliminal manipulation.

The European regulation, however, provides more detailed guidance on risk management throughout the lifecycle, requiring performance metrics, data quality, and comprehensive technical documentation. The Brazilian Bill, by contrast, relies on AIAs with public disclosure of their findings, embedding the precautionary principle and institutionalizing mechanisms for social participation and oversight, especially in cases where public authorities use high-risk AI.

In both cases, the effectiveness of such safeguards will depend on institutional capacity, technical resources, sectoral guidelines, and regular independent audits.

Regarding structural flaws, it is clear that the AI Act, although advanced, was designed primarily for special-purpose systems, and its adaptation to general-purpose AI - especially generative AI - remains dependent on delegated acts, evolving technical metrics (such as power thresholds) and multi-author coordination. This generates gray areas of responsibility in the deployment and integration chains.

Furthermore, the coexistence of *hard law* with codes of conduct and common specifications tends to fragment the regulatory landscape, deferring full normative coherence to a mosaic of frameworks still in development.

The Brazilian Bill, by contrast, reveals critical gaps in AI and generative AI models, as well as general-purpose systems, emotion recognition in asymmetric contexts (education and work), minimum standards for explainability and traceability, sanctions with real deterrent power for global agents, and the definition of the authority responsible for implementation of the legislation.

From the perspective of sufficiency, the overall balance reflects significant—but still conditional—normative advances.

At the European level, sufficiency will depend on the consolidation of the AI regulatory body, the maturation of harmonized standards and integration with existing regimes (data protection, services and digital markets). It will also require effective auditing and post-market monitoring practices capable of detecting and correcting biases, opacity and emerging behaviors in generative models.

In Brazil, sufficiency depends on defining and structuring the supervisory authority, converting impact assessment into real operational practice with clear methodologies, scopes, baselines, artifacts, and deadlines. It also requires enabling quality social participation and developing sectoral guidelines that translate principles into verifiable requirements—especially in the areas of

labor, credit, health, and justice, where power imbalances and potential harm are most severe.

In both cases, the interface with data protection, consumer protection, competition and sector regulation will be decisive in giving substantive meaning to the label, "risk-based".

Based on the diagnosis carried out and the weaknesses identified, it is possible to conclude that, firstly, duties applicable to general-purpose AI must be clarified throughout the chain - from the developer of fundamental models to the integrator and operator. These duties should include obligations of technical documentation (model and data cards), performance and bias testing, and duties of reciprocal cooperation to avoid a "vacuum of responsibility" in system integrations.

Second, minimum standards of explainability and accountability that are effectively operational and compatible with trade secrets (for example, counterfactual explanations and attribute importance metrics) must be defined, with publication in machine-readable formats and searchable repositories.

Third, regulation needs more consistent sanctioning guidelines to ensure a deterrent effect on large economic groups, including aggravating factors for recidivism and systemic impact.

Fourth, studies should be promoted to enable periodic independent audits in sensitive areas, in addition to impact assessment, with the publication of technical summaries and corrective-action plans.

Finally, the shortcomings identified in both frameworks do not invalidate their merits. Instead, they indicate that effective implementation will determine how real the level of protection of fundamental rights will be.

Sufficiency, therefore, does not stem solely from the legal text but it will undoubtedly be the result of the interaction between standards, institutions and, above all, the effective operationalization of practices.

REFERENCES

Beck, Ulrich. *Sociedade de Risco: Rumo a uma Outra Modernidade*. Editora 34 Ltda, 2010.

Bertoncini Correa, Cristina Mendes. Bonat, Débora. Rocha, Mariane Carolina Gomes da Silva. Braz, Fabricio Ataides. Silva, Nilton Correia da. Nishi, Luciana. Siqueira, Eduardo Camargo de. "Os desafios no desenvolvimento de inteligência artificial aplicada ao direito: uma análise sobre o projeto Osíris". *Revista Democracia Digital e Governo Eletrônico* 1, no. 24, Florianópolis (2025) 56-69

Bobbio, Norberto. *A Era dos Direitos*. Campus, 1992.

Bonat, Debora e Hartmann Peixoto, Fabiano, "GPTs e Direito: impactos prováveis das IAs generativas nas atividades jurídicas brasileiras"**,** *Seqüência: Estudos jurídicos e políticos* 44, (2023) 4-5.

Bower, Joseph L. and Clayton M. Christensen. Disruptive technologies: catching the wave. *Harvard Business Review* 73, no.1 (1995): 43–53. https://www3.yildiz.edu.tr/~naydin/MI2/lectures/Reading/Disruptive%20Technologie%20Catching%20the%20Wave.pdf.

Brasil. Senado Federal. *Projeto de Lei n. 2338/2023*. Dispõe sobre o uso da Inteligência Artificial. Senado Federal, 2024. https://www25.senado.leg.br/web/atividade/materias/-/materia/157233.

Canotilho, José Joaquim Gomes. *Direito Constitucional e Teoria da Constituição*. Almedina, 2000.

Chang, Heng *et al*. "A Survey of Large Language Models." *ACM Trans. Intell. Syst. Technol*. 15, no. 3, (2024): 1-45. https://dl.acm.org/doi/pdf/10.1145/3641289.

Data Science Academy. *Deep Learning Book*. 2025. https://www.deeplearningbook.com.br/introducao-as-redes-adversarias-generativas-gans-generative-adversarial-networks/.

Feigelson, Bruno. "A relação entre modelos disruptivos e o direito: estabelecendo uma análise metodológica baseada em três etapas." *Regulação e novas tecnologias*, coordenado por Rafael Veras Freitas, Leonardo Coelho Ribeiro e Bruno Feigelson. Fórum, 2018.

Giddens, Anthony. *Mundo em Descontrole*. Record, 2003.

Goodfellow, Ian *et al*. "Generative Adversarial Networks." *Advances in Neural Information Processing Systems* 3, no. 11 (2014): 1-9. https://www.researchgate.net/publication/263012109_Generative_Adversarial_Networks.

Hynek, Nik, Beata Guavurova e Matus Kubak. "Risks and benefits of artificial intelligence deepfakes: Systematic review and comparison of public attitudes in seven European Countries." *Journal of Innovation & Knowledge* 10, no. 5, (2025). https://www.sciencedirect.com/science/article/pii/S2444569X25001271.

Kaufman, Dora. *Desmistificando a Inteligência Artificial*. Autêntica, 2022.

Martens, Bruegel. "The European Union Ai Act: Premature or Precocious Regulation?" *Bruegel*, 2024. https://www.bruegel.org/analysis/european-union-ai-act-premature-or-precocious-regulation.

Sabrina Kutscher. "The EU AI Act: Law of Unintended Consequences?" *Technology and Regulation* 2025 (2025): 355-385. https://techreg.org/issue/view/1700.

Lecun, Yan, Geoffrey Hinton e Yoshua Bengio. "Deep Learning", *Nature* 521 (2015): 436–444, https://www.nature.com/articles/nature14539.

Lee, Kai-Fu. *Inteligência Artificial: como os robôs estão mudando o mundo, a forma como amamos, nos relacionamos, trabalhamos e vivemos*. Globo Livros, 2019.

Marmelstein, George. *Curso de direitos fundamentais*. Atlas, 2008.

McCulloch, Warren S. e Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* 52, no. 1/2 (1943): 99-115, https://www.cs.cmu.edu/~epxing/Class/10715/reading/McCulloch.and.Pitts.pdf.

ONU - Nações Unidas Brasil. *Objetivos do Desenvolvimento Sustentável: Trabalho decente e crescimento econômico*, 2025. https://brasil.un.org/pt-br/sdgs/8.

Saetra, Henrik Skaug. "Generative AI: Here to stay, but for good? *Technology in Society* 75 (2023). https://www.sciencedirect.com/science/article/pii/S0160791X2300177X.

Silva, José Benedito Lázaro da. "O efeito disruptivo das inovações e tecnologias frente às ciências jurídicas e sociais." *Regulação e novas tecnologias*, coordenado por Rafael Veras Freitas, Leonardo Coelho Ribeiro e Bruno Feigelson. Fórum, 2018.

Ribeiro, Leonardo Coelho. "A instrumentalidade do direito administrativo e a regulação de novas tecnologias disruptivas." *Regulação e novas tecnologias*, coordenado por Rafael Veras Freitas, Leonardo Coelho Ribeiro e Bruno Feigelson. Fórum, 2018.

Turing, Alan. Computing Machinery and Intelligence. *Mind* 59, no. 236 (1950): 433-460, https://turingtextos.blogspot.com/2017/02/computadores-e-inteligencia-alan.html.

Uniao Europeia. "Regulamento (UE) 2024/1689 do parlamento europeu e do conselho de 13 de junho de 2024." *Jornal da União Europeia*, 2024. https://eur-lex-europa-eu.translate.goog/legal-content/PT/TXT/?uri=CELEX:32024R1689&_x_tr_sl=en&_x_tr_tl=pt&_x_tr_hl=pt&_x_tr_pto=tc.

Yudkowsky, Eliezer. *Artificial Intelligence as a Positive and Negative Factor*. Oxford University Press, 2008. https://intelligence.org/files/AIPosNegFactor.pdf.

Zewe, Adam. Explicado: o Impacto Ambiental da IA Generativa. *MIT News on Campus and around the world*, 2025. https://news.mit.edu/2025/explained-generative-ai-environmental-impact-0117.

Received on 12/09/2025

Approved on 17/11/2025