REVISTA DE DIREITO INTERNACIONAL

BRAZILIAN JOURNAL OF INTERNATIONAL LAW

Editores responsáveis por essa edição:

Editores:

Nitish Monebhurrun

Ardyllis Alves Soares

Marcelo Dias Varella

Editores convidados:

David Ramiro Troitiño

Ignacio Bartesaghi

ISSN 2237-1036

Revista de Direito Internacional Brazilian Journal of International Law	Brasília	v. 20	n. 2	p. 1-633	ago	2023

doi: 10.5102/rdi.v20i2.9119

Artificial intelligence: a claim for strict liability for human rights violations*

IA e responsabilidade: uma abordagem internacional de direitos humanos

Lutiana Valadares Fernandes Barbosa**

Abstract

This article aims to answer the question of which civil responsibility threshold shall apply in case of violation of human rights in the context of artificial intelligence (AI). It starts by introducing possible risks AI presents to human rights, including privacy violations, discrimination, and lack of accountability. Next, it presents the challenges AI poses at the helm of civil responsibility such as challenges in attribution, break of the causal chain, and excessive burden of proof. It further discusses the under-development AI regulatory initiatives, such as the European Union´s and its expected Brussels effect, and the Brazilian framework, as regards responsibility and human rights, presenting their advantages and drawbacks. It then explores the alternative proposal of the principle of AI neutrality. It debates the AI responsibility threshold in the realm of human rights violations. The conclusion is that human rights are a necessary global filter for AI regulation and that strict liability is the most suitable threshold in the case of AI´s breaches of human rights.

Keywords: artificial intelligence; EU AI Act; responsibility; human rights; strict liability.

Resumo

Este artigo tem como objetivo refletir sobre inteligência artificial e responsabilidade civil a partir de uma abordagem de direitos humanos. Começa por introduzir possíveis riscos que a inteligência artificial (IA) apresenta para os direitos humanos. A seguir, apresenta os desafios que a IA coloca no comando da responsabilidade. Discute ainda as iniciativas regulatórias de IA em subdesenvolvimento, como a da União Europeia e seu esperado efeito Bruxelas, e o arcabouço brasileiro. Em seguida, reflete criticamente sobre as iniciativas regulatórias. Conclui que os direitos humanos são um filtro global necessário para a regulamentação da IA e para os pedidos de responsabilidade civil estrita no caso de violações dos direitos humanos por parte da IA.

Palavras chave: Inteligência Artificial, EU AI Act, Responsabilidade, Direitos Humanos, responsabilidade estrita.

* Recebido em 01/06/2023
Aprovado em 05/10/2023

** Defensora Pública Federal desde 2010. Mestre em Direito pela Columbia University/NY e PUC/MG. Doutora em Direito Internacional pela Universidade Federal de Minas Gerais.
E-mail: lutianafernandes@yahoo.com.br

1 Introduction

Artificial Intelligence (AI) is progressively ingraining itself into diverse aspects of human life. From the instant we wake up and engage with our smartphones, to the tasks we undertake at work, to the medical services we seek, to our entertainment choices, and even in our interactions with government services and benefits, AI integrates into our daily routines, often being unnoticed by users.

AI might mean accuracy and efficiency and lead to many positive outcomes, across diverse fields, such as improving medical care, by helping the diagnosis and treatment of diseases and improving education by aiding both teachers and students. However, studies have shown that AI might negatively impact human rights, in areas such as privacy violations and discrimination.

AI consists of systems that have the capability to perform tasks that are analogous or associated with tasks performed by humans, such as problem-solving, decision-making, language understanding, and learning. AI represents a paradigm shift in decision-making[1] and has been defined as a fourth industrial revolution[2]. Decision-making processes that previously could only be carried out by humans are increasingly carried out by autonomous devices operating by artificial intelligence. These devices recognize patterns, perform prediction and prediction often have learning capabilities.

Several legal challenges arise from this new reality, including accountability. Numerous AI systems operate as black boxes, which means a challenge to hold someone accountable for the violations. Adding to that AI--related decisions can be diffuse, a result of many smaller decisions of numerous people, making it difficult to ascribe responsibility. Therefore, careful consideration of responsibility for AI breaches of human rights is necessary aiming at thinking through measures to mitigate potential negative consequences. If there is a violation of human rights in the context of AI, who shall be held responsible? And what shall be necessary for responsibility to ensue?

This article focuses on civil liability in the context of the use of AI devices and analyzes it from an international, regional, and domestic law perspective, as regards human rights violations.

It aims to answer the question: which standard of responsibility shall apply in cases of human rights violations in the context of AI. It considers the existing and under-development initiatives and proposes, *de lege ferenda*, a responsibility threshold.

The analyses were conducted from a human rights--based approach.[3] To answer this question, qualitative research was developed. Bibliographical research was conducted regarding AI, Human Rights, and Responsibility. Its results were analyzed using the hypothetical--deductive method. A documentary survey was carried out in the under-development initiatives of AI regulation. Its results were analyzed using the inductive method.

[1] "Artificial Intelligence (AI) refers to systems or machines that mimic human cognitive functions such as learning, problem-solving, perception, decision-making, and natural language processing." - EUROPEAN COMMISSION. *What is Artificial Intelligence?* 2021. Avaiable in: https://ec.europa.eu/info/research-and-innovation/research-area/digital-transformation/artificial-intelligence_en. Accessed in: 13 nov. 2022. "Artificial Intelligence (AI) is the field of computer science dedicated to solving cognitive problems commonly associated with human intelligence, such as learning, problem-solving, and pattern recognition." - ARTIFICIAL Intelligence. *MIT Technology Review*, 2021. Avaiable in: https://www.technologyreview.com/topic/artificial-intelligence/. Accessed in: 13 nov. 2022. "Artificial Intelligence (AI) is the development of computer systems that can perform tasks that would require human intelligence to complete." - THE WORLD ECONOMIC FORUM. The World Economic Forum. *What is artificial intelligence?* We Forum, 2021. Avaiable in: https://www.weforum.org/agenda/2016/12/what-is-artificial-intelligence/. Accessed in: 13 nov. 2022. "Artificial Intelligence is the study of how to make computers do things which, at the moment, people do better" RUSSELL, S. J.; NORVIG, P. *Artificial intelligence*: A modern approach. Pearson Education, 2010. p. 2. Goodfellow, Bengio, and Courville (2016) in their book "Deep Learning": "AI is the science and engineering of making intelligent machines, especially intelligent computer programs" GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep learning*. MIT Press, 2016. p. 1.

[2] ABBOTT, Ryan. *The Reasonable Robot*: Artificial Intelligence and the Law, Cambridge University Press, 2020. p. 2 "Already impressive-sounding era titles such as the Fourth Industrial Revolution, the Second Machine Age, and the Automation Revolution are being used to describe the coming disruption."

[3] UNITED NATIONS SUSTAINABLE DEVELOPMENT GROUP. *Universal Values* - Principle One: Human Rights-Based Approach. UN, 2020. Avaiable in: https://unsdg.un.org/2030-agenda/universal-values/human-rights-based-approach. Accessed in: 13 nov. 2022. "The human rights-based approach (HRBA) is a conceptual framework for the process of human development that is normatively based on international human rights standards and operationally directed to promoting and protecting human rights. It seeks to analyses inequalities which lie at the heart of development problems and redress discriminatory practices and unjust distributions of power that impede development progress and often result in groups of people being left behind."

2 Some of the risks posed by AI at the helm of human rights

Human Rights protect all human beings, regardless of gender, national or ethnic origin, color, religion, age, or other characteristics. They are stated in the Universal Declaration of Human Rights, as well as many international treaties and customary international law. Human Rights embrace various rights such as life, education, health, freedom, and non-discrimination. Through ratifying treaties and customary international law, States are obliged to respect, protect, and remedy human rights violations within their jurisdiction.

AI might be beneficial in various domains but might also represent additional challenges to human rights.

In this sense, UNESCO´s recommendation on the Ethical Uses of AI states that:

> AI technologies can be of great service to humanity and all countries can benefit from them, but also raise fundamental ethical concerns, for instance regarding the biases they can embed and exacerbate, potentially resulting in discrimination, inequality, digital divides, exclusion and a threat to cultural, social and biological diversity and social or economic divides; the need for transparency and understandability of the workings of algorithms and the data with which they have been trained; and their potential impact on, including but not limited to, human dignity, human rights and fundamental freedoms, gender equality, democracy, social, economic, political and cultural processes, scientific and engineering practices, animal welfare, and the environment and ecosystems.[4]

In the field of privacy, AI-enabled surveillance systems, such as facial recognition, biometric data, and location tracking, can violate individual´s right to privacy by collecting and analyzing vast amounts of personal data without their consent. As regards discrimination, AI systems frequently learn from data, which may contain biases, enhancing, thus, discrimination against certain groups, such as people of African descent, women, migrants the elderly, and traditional peoples, in multiple areas such as hiring, lending, and criminal justice. In short, AI might be deployed with intended or unintended discriminatory results: racist, sexist, xenophobic, ageist, and ethnocidal.

Therefore, AI systems must pass through the Human Rights filter, which means they must Consider social risks, favor the underprivileged and vulnerable, consider the perspectives of indigenous people, riverside dwellers, quilombolas, women, mothers, the LGBTI population, be an instrument of social transformation and not the reinforcement of prejudices and discriminatory practices, ensure social and democratic participation, comply with the principles of non-discrimination, freedom of expression, privacy, among others.

3 AI and civil responsibility

In this section, the challenges AI poses to responsibility will be discussed. Next, the European and Brazilian regulatory initiatives will be presented as well as other doctrinal approaches. Then, their advantages and drawbacks and alternative proposals will be discussed.

3.1 Some of the Challenges posed by AI in the realm of responsibility

Schemes of responsibility are typically composed of tort, damage, causal link, and if the responsibility is not strict, the subjective element (fault or intent). However, in the context of AI, the subjective element might be very hard or even impossible to assess. AI might be capable of independent action, excluding the subjective element between the human decision to deploy it and the AI action performed or the decision taken. Furthermore, the subjective element might be inexistent due to AI devices' inherent unpredictability, which occurs both due to the complexity of the system's algorithms and to the interaction with the environment AI´s causal chain might be inaccessible even to programmers, so it is likely to be impossible to know and trace back how and why AI pursued the decision-making process.[5]

> Therefore current national liability rules, in particular, based on fault, are not suited to handling liability claims for damage caused by AI-enabled products and services. Under such rules, victims need to prove a wrongful action or omission by a person who caused the damage. The specific characteristics of AI, including complexity, autonomy and opacity

[4] UNESCO. *Recommendation on the Ethics of Artificial Intelligence*. Unesco, 2021. Avaiable in: https://unesdoc.unesco.org/ark:/48223/pf0000380455. Accessed in: 13 nov. 2022.

[5] ABBOTT, Ryan. *The Reasonable Robot*: Artificial Intelligence and the Law, Cambridge University Press, 2020. p. 33"An AI´s action cannot always be explained. It may be possible to determine what an AI has done, but not how or why it acted as it did."

> (the so-called "black box" effect), may make it difficult or prohibitively expensive for victims to identify the liable person and prove the requirements for a successful liability claim. In particular, when claiming compensation, victims could incur very high up-front costs and face significantly longer legal proceedings, compared to cases not involving AI. Victims may therefore be deterred from claiming compensation altogether.[6]

Furthermore, in the process of programming and deploying an AI system, there might be so many persons involved, them none of them can be considered responsible, as singularly considered, none of the actions suffice for the breach. That is named the "many hands problem" which creates challenges to the attribution of responsibility.

3.2 Regulatory Initiatives: EU and the Brussels effect

Regulatory initiatives are often classified as ""horizontal" or "vertical" regulations. "In a horizontal approach, regulators create one comprehensive regulation that covers the many impacts AI can have. In a vertical strategy, policymakers take a bespoke approach, creating different regulations to target different applications or types of AI."[7] The horizontal model bears the problem of lack of specificity regarding specific uses of AI, and might lack clarity as regards responsibility. On the other hand, the vertical model meaning a regulation for each use of AI, might create a regulatory chaos and hamper innovation, as each specific use has its specific regulations.

In Europe the EU AI project provides a horizontal framework, namely embracing a wide range of AI technologies. It is argued to be a positive strategy as it makes possible a harmonization of AI regulation across its member states. Some risks regarding this approach are that individual regulators who under the EU AI Act must enforce its requirements might interpret differently the same situation or might have varying capacities to regulate. It is also questioned if the proposed European AI Office will have the effective capability to complement national regulators to bring them to the same page. [8]Another challenge is that the proposed EU AI Act delegates technical requirements to standardization processes which is positive to allow some flexibility and expertise, but also bears the risk of lack of civil society participation, as it tends to be industry driven.[9]

The main feature of the project is the risk approach. AI uses presenting unacceptable risks must are prohibited, AI uses presenting high risk must pass through a conformity assessment, those offering limited risk have the duty of transparency, and those offering minimal risk have no extra AI requirements. Which technologies fall under which risk level is still being debated by EU. The risk approach was adopted aiming at leaving 80-90% of the uses of technology unregulated and thus not hampering innovation. Within the broad framework of the EU AI Act, EU is also discussing the EU AI Liability Directive[10].

In the explanatory memorandum to the EU AI Liability Directive, it was stated that:

> Current national liability rules, in particular based on fault, are not suited to handling liability claims for damage caused by AI-enabled products and services. Under such rules, victims need to prove a wrongful action or omission by a person who caused the damage. The specific characteristics of AI, including complexity, autonomy and opacity (the so-called "black box" effect), may make it difficult or prohibitively expensive for victims to identify the liable person and prove the requirements for a successful liability claim.[11]

[6] EUROPEAN COMMISSION. *Proposal for a directive of the european parliament and of the council on adapting non-contractual civil liability rules to artificial intelligence.* AI Liability Directive, 2022. Avaiable in: https://commission.europa.eu/system/files/2022-09/1_1_197605_prop_dir_ai_en.pdf. Accessed in: 13 nov. 2022.

[7] O'SHAUGHNESSY, Matt; SHEEHAN, Matt. Lessons From the World's Two Experiments in AI Governance. *Carnegier*, 14 feb. 2023. Avaiable in: https://carnegieendowment.org/2023/02/14/lessons-from-world-s-two-experiments-in-ai-governance-pub-89035. Accessed in: 13 nov. 2022.

[8] O'SHAUGHNESSY, Matt; SHEEHAN, Matt. Lessons From the World's Two Experiments in AI Governance. *Carnegier*, 14 feb. 2023. Avaiable in: https://carnegieendowment.org/2023/02/14/lessons-from-world-s-two-experiments-in-ai-governance-pub-89035. Accessed in: 13 nov. 2022.

[9] O'SHAUGHNESSY, Matt; SHEEHAN, Matt. Lessons From the World's Two Experiments in AI Governance. *Carnegier*, 14 feb. 2023. Avaiable in: https://carnegieendowment.org/2023/02/14/lessons-from-world-s-two-experiments-in-ai-governance-pub-89035. Accessed in: 13 nov. 2022.

[10] "The purpose of the AI Liability Directive proposal is to improve the functioning of the internal market by laying down uniform rules for certain aspects of non-contractual civil liability for damage caused with the involvement of AI systems." EUROPEAN COMMISSION. *Proposal for a directive of the european parliament and of the council on adapting non-contractual civil liability rules to artificial intelligence.* AI Liability Directive, 2022. Avaiable in: https://commission.europa.eu/system/files/2022-09/1_1_197605_prop_dir_ai_en.pdf. Accessed in: 13 nov. 2022.

[11] EUROPEAN COMMISSION. *Proposal for a directive of the european parliament and of the council on adapting non-contractual civil liability rules to*

Three possibilities were considered for the liability directive: To alleviate the victim´s burden of proof; To alleviate the victim´s burden of proof, to ensure strict liability for certain uses, and to ensure mandatory insurance; and to alleviate the victim´s burden of proof and after a period of time, to reevaluate the necessity of strict liability and mandatory insurance. The last option is the one conveyed by the EU liability directive, which is under consideration. Its article 3 deals with the disclosure of evidence and a rebuttable presumption of non-compliance, article 4 with a rebuttable presumption of a causal link in case of fault, and Article 5 with a review of the necessary measures after five years.[12]

The EU proposal, if adopted, will be the first regional regulation on AI. It is expected that it will impact the whole world, based on the so-called Brussels effect, which means the influence of EU regulations and standards on global governments, markets, and industries.[13]

On the positive side, the EU proposal as it currently stands has a great focus on human rights. Nonetheless, the liability directive proposal is based on fault. Despite alleviating the victim´s burden of proof, victims' protection is rather week, considering all the challenges posed by AI. Strict liability rules would be much better to ensure victims' redress. Therefore, on the accountability for AI violations of human rights, as it currently stands, the proposal is not a good influence on the world.

3.3 Regulatory Initiatives: Brazil

In Brazil, the proposals to regulate AI are also predominantly horizontal. The Bill n. 21/2020 aimed at establishing "fundamentals, principles, and guidelines for the development and application of artificial intelligence." It states that Human Rights is one of the fundamentals of the development and use of AI in Brazil (art.4o). Initially it stated that the general rule shall be subjective responsibility (fault or intent, unlawful act, damage, causal link) in the context of AI violations. Strict liability was foreseen just for consumer relations and government responsibility. Its responsibility provisions were highly criticized. Suggestions regarding how to address responsibility varied broadly. From strict liability to schemes of responsibility. It remains to be seen how the bill will be enacted.

In 2023, after the thoughtful work of a commission of lawyers, Bill N. 21/2020 was substituted to Bill n. 2338/2023. The project is grounded on human rights. It states that the development, implementation, and use of systems of AI in Brazil are grounded on "the centrality of the human person" and on the "respect for human rights and democratic values."[14] Regarding responsibility, it is more protective than the EU proposal, as it states that:

> Art. 27. The supplier or operator of the system of artificial intelligence that causes patrimonial, moral, individual, or collective damage is obligated to repair it fully, regardless of the degree of autonomy of the system.
>
> § 1 In the case of a system of artificial intelligence of high risk or excessive risk, the supplier or operator responds strictly for the damage caused, to the extent of its participation in the damage.
>
> § 2 When it is not a system of artificial intelligence of high risk, the fault of the agent causing the damage will be presumed, applying the reversal of the burden of proof in favor of the victim.[15]

It is positive that under the Brazilian Bill, n. 2338/2023 strict liability ensues in case of systems of AI of high or excessive risk. This means it offers a higher protection threshold compared to the EU project,

artificial intelligence. AI Liability Directive, 2022. Avaiable in: https://commission.europa.eu/system/files/2022-09/1_1_197605_prop_dir_ai_en.pdf. Accessed in: 13 nov. 2022.

[12] "Three policy options were assessed: Policy option 1: three measures to ease the burden of proof for victims trying to prove their liability claim. Policy option 2: the measures under option 1 + harmonising strict liability rules for AI use cases with a particular risk profile, coupled with a mandatory insurance. Policy option 3: a staged approach consisting of: – a first stage: the measures under option 1; – a second stage: a review mechanism to re-assess, in particular, the need for harmonising strict liability for AI use cases with a particular risk profile (possibly coupled with a mandatory insurance)." https://commission.europa.eu/business-economy-euro/doing-business-eu/contract-rules/digital-contracts/liability-rules-artificial-intelligence_en. EUROPEAN COMMISSION. *Proposal for a directive of the european parliament and of the council on adapting non-contractual civil liability rules to artificial intelligence*. AI Liability Directive, 2022. Avaiable in: https://commission.europa.eu/system/files/2022-09/1_1_197605_prop_dir_ai_en.pdf. Accessed in: 13 nov. 2022.

[13] For an In-depth knowledge on the Brussels effect see BRADFORD, Anu. *The Brussels Effect*: How the European Union Rules the World. Oxford University Press, 2020.

[14] PACHECO, Rodrigo. *PL 2338/2023*. Dispõe sobre o uso da Inteligência Artificial. Avaiable in: https://www25.senado.leg.br/web/atividade/materias/-/materia/157233. Accessed in: 13 nov. 2022. art. 2º

[15] PACHECO, Rodrigo. *PL 2338/2023*. Dispõe sobre o uso da Inteligência Artificial. Avaiable in: https://www25.senado.leg.br/web/atividade/materias/-/materia/157233. Accessed in: 13 nov. 2022.

where there is only a provision to alleviate the victim´s burden of proof. However, strict liability only ensues for AI of high risk or excessive risk, and AI-related actions that impact human rights are part of the high or excessive list. For the non-high-risk AI, there is a reversal of the burden of proof, which still helps victims to receive compensation but is not as protective as strict liability.

Hopefully, the Brazilian bill will be amended to embrace AI-related actions that impact human rights as part of the high or excessive list. Optimistically, the Brazilian bill will influence other Inter-American member states and future Inter-American regulation. Unfortunately, power dynamics are so that global south legislations have a low or null impact on EU legislation, meaning that the chances that the Brazilian bill is not likely to influence the EU bill.

3.4 The principle of AI neutrality approach.

As can be observed from the Brazilian and EU proposals, some states are discussing strict liability in the context of AI, [16] as well as other approaches focused on providing redress to victims such as reversal or facilitation of the burden of proof.

On the doctrine, there are other approaches, such as the principle of AI neutrality coined by Abbot. The principle of AI neutrality, states that strict liability should not be applied if AI is safer or performs better than humans and humans are not subject to strict liability.[17] This means that despite recognizing the benefits of strict liability, he claims that if AI offers better and safer results, for example, in driving a car, it is not effective that human´s liability is grounded on fault and AI is subjected to strict liability, among his arguments is that a higher threshold of accountability could hamper innovation.

Abbot´s proposal has the positive of considering the importance of innovation. However, Abbot does not analyze responsibility with a human rights approach, and in the specific context of breaches of human rights.

Our claim is that the principle of AI neutrality shall not be applied in cases of human rights-related AI breaches, as they relate to the most essential rights of human beings and shall be subjected to strict liability as will be exposed in the next section. Requiring the scrutiny of intentionality (intent, fault, negligence, recklessness) for AI actions is far more challenging compared to human actions, considering AI´s inherent unpredictability, which might create an unsurmountable barrier to remedying human rights violations. From a human rights-centric perspective higher threshold of responsibility does not hamper innovation, but rather creates opportunities for a human rights-aligned innovation that by design cares more about preventing possible damages.

4 Strict liability for human rights breaches

First the EU and the Brazilian proposals for regulation of AI bear in common the human-centric and human rights-grounded perspectives. In the same sense, initiatives such as the Asilomar Principles and the UNESCO Recommendation on the Ethics of AI are also grounded on human rights and responsibility.

According to the Asilomar Principle 9, named **Responsibility, "**Designers and builders of advanced AI systems are stakeholders in the moral implications of their use, misuse, and actions, with a responsibility and opportunity to shape those implications."[18] Asilomar Principle 11 is named **Human Values, and states that** "AI systems should be designed and operated so as to be compatible with ideals of human dignity, rights, freedoms, and cultural diversity." [19]

UNESCO Ethical Uses of AI is grounded on the respect of human rights and responsibility.

> 42. AI actors and Member States should respect, protect and promote human rights and fundamental freedoms, and should also promote the protec-

[16] GEIß, Robin (org.). *Lethal Autonomous Weapons Systems*: Risk Management and State Responsibility in Lethal Autonomous Weapons Systems Technology, Definition, Ethics, Law & Security. Berlin: German Federal Foreign Office. Avaiable in: https://www.auswaertiges-amt.de/blob/204830/5f26c2e0826db0d000072441fdeaa8ba/abruestung-laws-data.pdf. Accessed in: 17 nov. 2022. p. 117

[17] ABBOTT, Ryan. *The Reasonable Robot*: Artificial Intelligence and the Law, Cambridge University Press, 2020. p. 4.

[18] FUTURE OF LIFE INSTITUTE. *Asilomar Principles*. 2017. Avaiable in: https://futureoflife.org/open-letter/ai-principles/. Accessed in: 13 nov. 2022.

[19] FUTURE OF LIFE INSTITUTE. *Asilomar Principles*. 2017. Avaiable in: https://futureoflife.org/open-letter/ai-principles/. Accessed in: 13 nov. 2022.

> tion of the environment and ecosystems, assuming their respective ethical and legal responsibility, in accordance with national and international law, in particular Member States' human rights obligations, and ethical guidance throughout the life cycle of AI systems, including with respect to AI actors within their effective territory and control. The ethical responsibility and liability for the decisions and actions based in any way on an AI system should always ultimately be attributable to AI actors corresponding to their role in the life cycle of the AI system.[20]

Despite the Asilomar Principles, the UNESCO recommendation, and the EU and the Brazilian proposals of bills grounds on responsibility and human rights, none of them are explicit to facilitate responsibility with a focus on providing redress to victims when human rights are violated. The Brazilian proposal goes much further than the EU and ensures strict liability for systems of AI of high or excessive risk but does not include among the AI of high or excessive risk specifically Human Rights violations.

This paper claims that if the damage caused was a violation of human rights, strict liability shall ensue, grounded on the following arguments.

First, there is an obligation to respect, protect and remedy human rights violations. Therefore, human rights obligations must be observed in all domains and are thus a necessary filter for the development and use of AI, meaning that they can only be developed and used in conformity with human rights. In line with this, the main AI principles, such as the aforementioned Asilomar Principles, the UNESCO recommendation on the Ethical Uses of AI, the under-development EU AI Act, and the Brazilian bill are all grounded on human rights and foster the obligations to protect and respect human rights but do not address the obligation to effectively remedy.

Second, AI is inherently unpredictable, meaning it might perform in venues unforeseen even by programmers, creating additional risks to human rights. These additional risks, combined with the obligation to remedy human rights violations, are a basis for strict liability.

The report from the EU Parliament Commission report[21] that accompanied the EU White Paper on AI, highlights AI characteristics that lead to unpredictability and create obstacles to comprehending the root causes of damages. "They can combine connectivity, autonomy, and data dependency to perform tasks with little or no human control or supervision."[22] Moreover, AI systems are characterized by a "[…] plurality of economic operators involved in the supply chain and the multiplicity of components, parts, software, systems or services, which together form the new technological ecosystems."[23] They can also learn from data, experiences, and self-update. This means that "The vast amounts of data involved, the reliance on algorithms and the opacity of AI decision-making, make it more difficult to predict the behavior of an AI-enabled product and to understand the potential causes of a damage."[24]

The inherent unpredictability of AI system means the impossibility or challenges to predict precisely what courses an AI system will pursue to achieve its objectives, even if the high-level goals are set.

> Unpredictability of AI, one of many impossibility results in AI Safety, also known as Unknowability [Vinge, 1993] or Cognitive Uncontainability [Cognitive Uncontainability, 287019], is deemed as our inability to precisely and consistently predict what specific actions an intelligent system will take to achieve its objectives, even if we know the terminal goals of the system. It is related but is not the same as unexplainability and incomprehensibility of AI [Yampolskiy, 2019]. Unpredictability does not imply that better-than-random statistical analysis is impossible; it simply points out a general limitation on how well such efforts can perform, and is parti-

[20] UNESCO. *Recommendation on the Ethics of Artificial Intelligence*. Unesco, 2021. Avaiable in: https://unesdoc.unesco.org/ark:/48223/pf0000380455. Accessed in: 13 nov. 2022.

[21] COMMISSION TO THE EUROPEAN PARLIAMENT; EUROPEAN ECONOMIC AND SOCIAL COMMITTEE. *Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics COM/2020/64*. Avaiable in: https://eur-lex.europa.eu/legal-content/en/TXT/?qid=1593079180383&uri=CELEX:52020DC0064. Accessed in: 13 nov. 2022.

[22] COMMISSION TO THE EUROPEAN PARLIAMENT; EUROPEAN ECONOMIC AND SOCIAL COMMITTEE. *Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics COM/2020/64*. Avaiable in: https://eur-lex.europa.eu/legal-content/en/TXT/?qid=1593079180383&uri=CELEX:52020DC0064. Accessed in: 13 nov. 2022.

[23] COMMISSION TO THE EUROPEAN PARLIAMENT; EUROPEAN ECONOMIC AND SOCIAL COMMITTEE. *Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics COM/2020/64*. Avaiable in: https://eur-lex.europa.eu/legal-content/en/TXT/?qid=1593079180383&uri=CELEX:52020DC0064. Accessed in: 13 nov. 2022.

[24] COMMISSION TO THE EUROPEAN PARLIAMENT; EUROPEAN ECONOMIC AND SOCIAL COMMITTEE. *Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics COM/2020/64*. Avaiable in: https://eur-lex.europa.eu/legal-content/en/TXT/?qid=1593079180383&uri=CELEX:52020DC0064. Accessed in: 13 nov. 2022.

cularly pronounced with advanced generally intelligent systems (superintelligence) in novel domains.[25]

For example, Deep Blue AI chess player developers cannot foresee every move it will take. What they can anticipate is that it moves to aim at winning.[26]

Under the threshold of strict liability, damage gives rise to responsibility, and the issue of intentionality (intent, fault, negligence, recklessness) is removed from scrutiny. Agents are automatically responsible when AI causes damage, even if it behaves unpredictably. Requiring the scrutiny of the issue of intentionality (intent, fault, negligence, recklessness) for AI misdoings is far more challenging compared to human misdoings, considering AI´s is capable to perform activities that previously only humans could develop and that it is inherently unpredictable both due to the interaction with the environment and due to the complexity and black boxes of algorithms. A threshold other than strict liability might create an unsurmountable barrier to remedying human rights violations.

Third, strict liability for human rights violations incentivizes a human rights-centric perspective through the design of AI programs and devices. UNESCO´s hallmark instrument on the ethical uses of AI states that "risks and ethical concerns should not hamper innovation and development but rather provide new opportunities and stimulate ethically-conducted research and innovation that anchor AI technologies in human rights and fundamental freedoms, values and principles, and moral and ethical reflection."[27] If developers and deployers are aware of a strict liability threshold if they violate human rights, when they develop, test, purchase, and deploy AI devices, they will have e extra precautions concerns with human rights risk mitigating measures.

Fourth, at the national and international levels, there is a tendency towards strict liability regimes regarding the development of dangerous activities.[28] Within the domestic domain, some States discuss strict liability for autonomous technologies.[29] Internationally, environmental law has deep-rooted strict liability regimes for hazardous activities with principles such as the "polluter-pays."[30] The Convention on Civil Liability for Damage Resulting from Activities Dangerous to the Environment, for instance, states at its preamble the "desirability of providing for strict liability in this field taking into account the "Polluter Pays" Principle."[31] The International Convention on Civil Liability for Oil Pollution Damage (CLC) foresees for example, that "the owner of a ship at the time of an incident, or where the incident consists of a series of occurrences at the time of the first such occurrence, shall be liable for any pollution damage caused by oil which has escaped or been discharged from the ship as a result of the incident."[32] The Convention on International Liability for Damage Caused by Space Objects also foresees strict liability in its article 2o "A launching State shall be absolutely liable to pay compensation for damage caused by its space object on the surface of the earth or to aircraft flight."[33]

The aforementioned instruments point to the importance of strict liability of risky activities.

5 Conclusion

AI and its ethical and legal implications must be seen through the lenses of human rights, the public interest,

[25] YAMPOLSKIY, Roman. Unpredictability of AI: On the Impossibility of Accurately Predicting All Actions of a Smarter Agent. *Journal of Artificial Intelligence and Consciousness*, v. 7, 2020. p. 110.

[26] YAMPOLSKIY, Roman. Unpredictability of AI: On the Impossibility of Accurately Predicting All Actions of a Smarter Agent. *Journal of Artificial Intelligence and Consciousness*, v. 7, 2020. p. 110.

[27] UNESCO. *Recommendation on the Ethics of Artificial Intelligence*. Unesco, 2021. Avaiable in: https://unesdoc.unesco.org/ark:/48223/pf0000380455. Accessed in: 13 nov. 2022.

[28] AMOROSO, Daniele. *Autonomous Weapons Systems and International Law*: A Study on Human-Machine Interactions in Ethically and Legally Sensitive Domains. Edizioni Scientifiche Italiane, 2020.

[29] GEIß, Robin (org.). *Lethal Autonomous Weapons Systems*: Risk Management and State Responsibility in Lethal Autonomous Weapons Systems Technology, Definition, Ethics, Law & Security. Berlin: German Federal Foreign Office. Avaiable in: https://www.auswaertiges-amt.de/blob/204830/5f26c2e0826db0d000072441fdeaa8ba/abruestung-laws-data.pdf. Accessed in: 17 nov. 2022. p. 117

[30] UNITED NATIONS. *General Assembly Report of the United Nations Conference on Environment and Development*. Rio de Janeiro, 3-14 June 1992. v. 1. Annex I Rio declaration on environment and development A/CONF.151/26 (Vol. I) (12 August 1992) At article 16

[31] CONCIL OF EUROPE. *Convention on Civil Liability for Damage Resulting from Activities Dangerous to the Environment*. Adopted 12 June 1993, entry into force 12 June 1993. Lugano, 21, v. 6, 1993. Avaiable in: https://rm.coe.int/168007c079. Accessed in: 13 nov. 2022.

[32] INTERNATIONAL MARITIME ORGANIZATION. *Convention on Civil Liability for Oil Pollution Damage*. UN, 1992. Avaiable in: https://treaties.un.org/doc/Publication/UNTS/Volume%20973/volume-973-I-14097-English.pdf. Accessed in: 17 nov. 2022.

[33] UNITED NATIONS. *Convention on International Liability for Damage Caused by Space Objects*. UN, 29 mar. 1972. This Convention elaborates on article 7 of the UNITED NATIONS. *Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, including the Moon and Other Celestial Bodies*. 27 jan. 1967. art. 7.

and the United Nations development goals. These Human Rights lenses shall apply to all documents aiming at regulating AI, including the upcoming EU AI Act, the under-development Brazilian Regulation on AI, as well as to future Inter-American instruments on AI, which must observe the international responsibility to protect respect and remedy human rights violations. In this context, this paper claims strict liability is a relevant instrument to adequately provide remedies for AI-related human rights violations in the context of AI breaches and promote justice for victims.

Four main arguments were presented as the lynchpin for strict liability for human rights breaches in the context of AI. First, considering that human rights are essential rights that belong to every person in the world and the obligation to protect, respect, and remedy human rights violations, and the fact that the main existing AI instruments are based on a human rights approach, but do not provide adequate remedies for human rights violations. Second, AI develops activities formerly just performed by human beings and creates additional risks due to its inherent unpredictability and black boxes. Third, the necessity to foster human rights by design and risk mitigation approach in the context of the development, purchase, and deployment of AI devices. Fourth, the trend of national and international instruments for strict liability in the context of risky activities, such as in the human right to environment context.

In case AI systems cause damages that violate human rights, as argued throughout the paper, strict liability shall ensue. Ideally, this approach shall be adopted internationally, as AI violations often cross borders and are not limited to one jurisdiction. The under-development Brazilian project of bill PL 2338/2023 steps in this direction, while not yet ensuring strict liability for all human rights-related breaches. The EU project is still far from this perspective but shall also follow this path to comply with international obligations to respect protect, and remedy human rights violations.

References

ABBOTT, Ryan. *The Reasonable Robot*: Artificial Intelligence and the Law, Cambridge University Press, 2020.

AMOROSO, Daniele. *Autonomous Weapons Systems and International Law*: A Study on Human-Machine Interactions in Ethically and Legally Sensitive Domains. Edizioni Scientifiche Italiane, 2020.

ARTIFICIAL Intelligence. *MIT Technology Review*, 2021. Avaiable in: https://www.technologyreview.com/topic/artificial-intelligence/. Accessed in: 13 nov. 2022.

BRADFORD, Anu. *The Brussels Effect*: How the European Union Rules the World. Oxford University Press, 2020.

COMMISSION TO THE EUROPEAN PARLIAMENT; EUROPEAN ECONOMIC AND SOCIAL COMMITTEE. *Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics COM/2020/64*. Avaiable in: https://eur-lex.europa.eu/legal-content/en/TXT/?qid=1593079180383&uri=CELEX:52020DC0064. Accessed in: 13 nov. 2022.

CONCIL OF EUROPE. *Convention on Civil Liability for Damage Resulting from Activities Dangerous to the Environment*. Adopted 12 June 1993, entry into force 12 June 1993. Lugano, 21, v. 6, 1993. Avaiable in: https://rm.coe.int/168007c079. Accessed in: 13 nov. 2022.

EUROPEAN COMMISSION. *Proposal for a directive of the european parliament and of the council on adapting non-contractual civil liability rules to artificial intelligence*. AI Liability Directive, 2022. Avaiable in: https://commission.europa.eu/system/files/2022-09/1_1_197605_prop_dir_ai_en.pdf. Accessed in: 13 nov. 2022.

EUROPEAN COMMISSION. *What is Artificial Intelligence?* 2021. Avaiable in: https://ec.europa.eu/info/research-and-innovation/research-area/digital-transformation/artificial-intelligence_en. Accessed in: 13 nov. 2022.

FUTURE OF LIFE INSTITUTE. *Asilomar Principles*. 2017. Avaiable in: https://futureoflife.org/open-letter/ai-principles/. Accessed in: 13 nov. 2022.

GEIß, Robin (org.). *Lethal Autonomous Weapons Systems*: Risk Management and State Responsibility in Lethal Autonomous Weapons Systems Technology, Definition, Ethics, Law & Security. Berlin: German Federal Foreign Office. Avaiable in: https://www.auswaertiges-amt.de/blob/204830/5f26c2e0826db0d000072441fdeaa8ba/abruestung-laws-data.pdf. Accessed in: 17 nov. 2022.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep learning*. MIT Press, 2016.

INTERNATIONAL MARITIME ORGANIZATION. *Convention on Civil Liability for Oil Pollution Damage*. UN, 1992. Avaiable in: https://treaties.un.org/doc/Publication/UNTS/Volume%20973/volume-973-I-14097-English.pdf. Accessed in: 17 nov. 2022.

O'SHAUGHNESSY, Matt; SHEEHAN, Matt. Lessons From the World's Two Experiments in AI Governance. *Carnegier*, 14 feb. 2023. Avaiable in: https://carnegieendowment.org/2023/02/14/lessons-from-world-s-two-experiments-in-ai-governance-pub-89035. Accessed in: 13 nov. 2022.

PACHECO, Rodrigo. *PL 2338/2023*. Dispõe sobre o uso da Inteligência Artificial. Avaiable in: https://www25.senado.leg.br/web/atividade/materias/-/materia/157233. Accessed in: 13 nov. 2022.

RUSSELL, S. J.; NORVIG, P. *Artificial intelligence*: A modern approach. Pearson Education, 2010.

THE WORLD ECONOMIC FORUM. The World Economic Forum. *What is artificial intelligence?* We Forum, 2021. Avaiable in: https://www.weforum.org/agenda/2016/12/what-is-artificial-intelligence/. Accessed in: 13 nov. 2022.

UNESCO. *Recommendation on the Ethics of Artificial Intelligence*. Unesco, 2021. Avaiable in: https://unesdoc.unesco.org/ark:/48223/pf0000380455. Accessed in: 13 nov. 2022.

UNITED NATIONS SUSTAINABLE DEVELOPMENT GROUP. *Universal Values* - Principle One: Human Rights-Based Approach. UN, 2020. Avaiable in: https://unsdg.un.org/2030-agenda/universal-values/human-rights-based-approach. Accessed in: 13 nov. 2022.

UNITED NATIONS. *Convention on International Liability for Damage Caused by Space Objects*. UN, 29 mar. 1972.

UNITED NATIONS. *General Assembly Report of the United Nations Conference on Environment and Development*. Rio de Janeiro, 3-14 June 1992. v. 1.

UNITED NATIONS. *Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, including the Moon and Other Celestial Bodies*. 27 jan. 1967.

YAMPOLSKIY, Roman. Unpredictability of AI: On the Impossibility of Accurately Predicting All Actions of a Smarter Agent. *Journal of Artificial Intelligence and Consciousness*, v. 7, 2020.