BRAZILIAN STRATEGY OF ARTIFICIAL INTELLIGENCE (EBIA) AND PUBLIC POLICIES: PROPOSALS FOR THE IMPLEMENTATION OF THE LEGISLATION, REGULATION, AND ETHICAL USE AXES AND GOVERNANCE OF AI

Sthéfano Bruno Santos Divino[1]

Abstract: This article has as its research problem the following question: how and which public policies can be appropriate for the effectiveness and implementation of the Brazilian Strategy for Artificial Intelligence (EBIA) regarding the axes Legislation, Regulation, and Ethical Use, and Governance of Artificial Intelligence (AI)? The aim is to present the EBIA and its axes of implementation and, afterward, to contextualize it according to Brazil's classification in terms of Artificial Intelligence in the international scenario. Global Innovation Index prepared by the World Intellectual Property Organization and the Artificial Intelligence Index Report prepared by Stanford University are used for this purpose. The methodology used for this is monographic research. In the end, it is concluded that the EBIA is a mechanism capable of inserting Brazil into the AI regulatory race, but it needs strong and effective public policies aimed at its implementation.

Keywords: Brazilian Artificial Intelligence Strategy; governance; artificial intelligence; public policy; regulation.

1 Introduction: The Brazilian Artificial Intelligence Strategy (EBIA in Portuguese) and its implementation axes

Artificial Intelligence[2] (AI) has rapidly evolved and developed to offer economic and social benefits. In recent years, a need has arisen for adopting regulatory measures and public policies[3] by the Executive Branch for implementing AI in both the domestic and industrial segments. Canada (2017), China (2017), Denmark (2021), the European Commission (2021), Finland (2017), France (2018), India (2018), Italy (2020), Japan (2018), Mexico (2018), the Nordic Countries (2018), Singapore (2020), South Korea (2019), Sweden (2018), Taiwan (2018), and the UK (2020) have launched strategies to promote the use and development of AI (Dutton, 2018). While none of the strategies are similar, we focus on different aspects of AI: scientific research, talent development and capture, skills and education, adoption in the public and private sectors, ethics and inclusion, regulatory standards, and digital infrastructure (DUTTON, 2018).

---

[1] Doctorate student (2020 - Scholarship holder of the Academic Excellence Program - Proex - Capes/Taxa) and Master (2019) in Private Law from the Pontifícia Universidade Católica de Minas Gerais. Bachelor of Law from the Centro Universitário de Lavras (2017). Associate Professor of the Law Program at the Centro Universitário de Lavras (2020 - present). Substitute Professor of Private Law at the Universidade Federal de Lavras (03/2019 - 03/2021). Lawyer. Orcid iD: https://orcid.org/0000-0002-9037-0405. Lattes: http://lattes.cnpq.br/5133514180104561. Email: sthefanoadv@hotmail.com.

[2] One of the most accepted definitions in the scientific branch – although also criticized – is that of Russell and Norvig, who define Artificial Intelligence as "[...] the study of agents that receive percepts from the environment and perform actions". (RUSSELL; NORVIG, 2010, p. VIII).

[3] About definitions and history of public policies, see more at (SOUZA, 2002)

Also, at the international level, on November 22nd, 2021, the chair of the United Nations Educational, Scientific and Cultural Organization (UNESCO, 2021) launched the first international agreement on Ethics in AI applications. The recommendations also address data protection, the prohibition of the social practice of score (CITRON; PASQUALE, 2014; SERASA, 2021), and mass surveillance to monitor and evaluate systems during their implementation and execution, and environmental protection to the possibility of an AI to use energy and other primary services more efficiently.

Recognizing that the significant increase in computational power through practical advances in machine learning[4] enables successes in a range of applied domains, drawing attention to public policy and business development to join the race for global leadership in AI, Brazil (2021) launches its Brazilian Artificial Intelligence Strategy (EBIA)[5] aiming its performance in fields such as labor, education, taxation, research, development, and innovation and ethics.

EBIA (BRASIL, 2021) intends to guide the role of state actions for the development of actions, whatever their aspects, to stimulate research, innovation, and development of AI solutions and ensure that their use is conscious, ethical, legal, and for the benefit of a better future. EBIA was built in three stages:

1) Hiring specialized AI consulting;
2) National and international benchmarking;
3) Public consultation process.

Specialized consultancy was hired by the Ministry of Science, Technology, Innovations, and Communications (MCTI) through the International Technical Cooperation Project (PRODOC in Portuguese) with UNESCO. On the other hand, the public consultation was carried out through the Federal Government's electronic platform between December 12th, 2019, and March 3rd, 2020, a period in which about a thousand contributions were received and used as the basis for the construction of the EBIA (BRASIL, 2021).

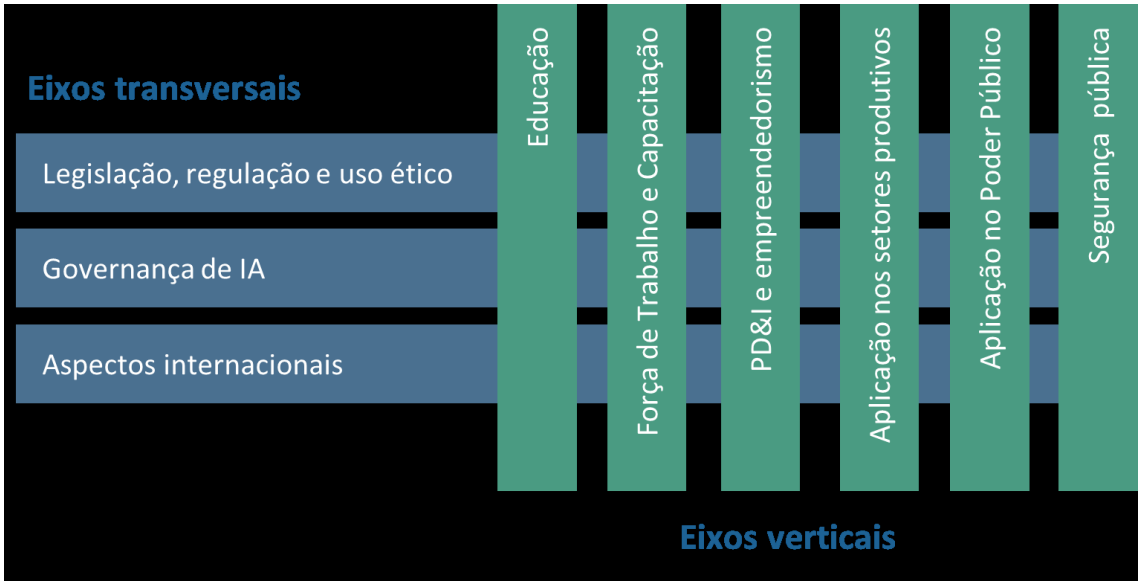
EBIA follows the recommendations of the Organization for Economic Cooperation and Development (OECD, 2019) on AI, highlighting, among others: a) its benefit to people and the planet; b) respect for the Rule of Law, human rights, democratic values, and diversity; c) transparency aimed at general understanding about AI systems; and d) robust functioning throughout its life cycle.

Therefore, EBIA works with nine thematic axes, three of which are transversal axes - 1)

[4] "Machine learning is an evolving branch of computational algorithms that are designed to emulate human intelligence by learning from the surrounding environment. They are considered the working horse in the new era of the so-called big data" (EL NAQA; MURPHY, 2015. p. 3-11).
[5] Established by MCTI Ordinance nº 4,617 of April 6th, 2021, amended by MCTI Ordinance nº 4,979 of July 13th, 2021 (Annex).

Legislation, regulation, and ethical use; 2) AI governance[6]; 3) International aspects -, and six vertical axes - 1) Education; 2) Workforce and training; 3) PDI and entrepreneurship; 4) Expansion in the Public Power; 5) Application in the productive sectors; and 6) Public Security. The transverse axes act as the basis for constructing the other vertical sectors. The assumptions of the latter must be present for all (see Figure 1). Although they appear to be independent since the subjects are diverse and specifically cover each sector, they should all be abstract enough to embrace all AI sectors in the scope of creating the transversal axes.

Figure 1 – EBIA axes



Source: Brasil (2021).
Note: Transversal axes: Legislation, regulation, and ethical use; Ai governance; International aspects. Vertical axes: Education; Workforce and Training; R&D and entrepreneurship; Application in productive sectors; Application in the Public Power; Public safety.

In this sense, it can be said that the algorithms[7] must be neutral (*agnostics*) and not focus specifically on an AI modality/species or data analysis methodology. The strategy must be applied in the constitution, application, and use of any and all AI. In other words, technology is neutral, should not focus on systems, software or specific techniques, and should be applied regardless of the development of computational language and data storage techniques. The

[6] For Floridi, "digital governance is the practice of establishing and implementing policies, procedures, and standards for the proper development, use, and management of *infosphere*. Through human supervision, it is intended to ensure that an AI system does not compromise human autonomy or produce negative effects. Digital governance can include guidelines and recommendations that overlap with, but are not identical to, digital regulation. This is just another form of considering the relevant legislation, a system of laws drawn up and applied through social or government institutions to regulate the behavior of relevant agents in the *infosphere*" (FLORIDI L. 2018).

[7] "Algorithms are the basis of the *software* development process and are part of the tools by which programmers create strategies to fractionate problems into steps and processes that can be computationally translated. There are examples of all levels of complexity in technology. The computer startup process a simpler application of algorithm: there is a *software* - basically the computer translation of an algorithm - in charge of testing all the components of your computer to know if everything is in order and, after that, look for the operating system on the disk to load it" (GARRET, 2020, our translation).

adoption of these practices serves as baselines for a series of considerations and measures for organizations to operate in any industry and concretely adopt the strategy. Specific industries or organizations may adopt additional considerations and technical measures to adapt their production line according to their needs. Thus, the strategy should not focus on public or private organizations according to their size or constitutive modality.[8]

The proposal making these claims is outlined through objectives. The EBIA aims to contribute to developing ethical principles for developing and using more responsible AI. At the same time, the balance between sustained investments in AI research and development is assumed to remove barriers to AI innovation. It is intended to train professionals for the AI ecosystem to stimulate innovation and the development of Brazilian AI in an international environment. Thus, cooperation can be promoted between national and international and public and private entities, and industry and research centers for the development of AI.

Such objectives are not easy to achieve in the current conjuncture. This article has the following question as a **search problem**: how and what public policies can be adjusted for the effectiveness and implementation of EBIA regarding the legislation, regulation, and ethical use axes and governance of AI?

After contextualization, the first section analyzes Brazil's classification in terms of AI in the international scenario. The Global Innovation Index, prepared by the World Intellectual Property Organization, and the Artificial Intelligence Index Report, prepared by Stanford University, were used. At this point, quantitative research is carried out on the sites Conecta Startup and Startup Brasil to verify the adequacy of international indices to the degree of innovation in the national territory. As a result, out of 268 startups, the following sectors predominate: agribusiness (14), industrial (10), health and welfare (15), education (23), health (17), IT and telecommunications (19), retail (13), finance (11), and media and communication (12).

Original contributions show that public policies aimed at scientific advancement in Information Technology and, specifically, AI, are not in accordance with the claims of EBIA. It is worth noting one limitation of this study. As pointed out, the qualitative research demonstrates performance sectors such as IT, telecom, and hardware. In this sense, it is unfeasible to verify whether the startup members of these branches work specifically in the AI sector. In other words, the result obtained can be more negative if verified concretely and in detail.

Additionally, the second section is responsible for contributions to the legislation, regulation, and ethical use axis, especially in the thematic of civil liability of practical acts by AI. The main result postulates that the adequate responsibility for the illicit acts practiced by

[8] Reviews excerpted from SINGAPORE (2020).

artificially-intelligent entities is the fault liability since it allows the reduction of the indemnifying duty based on the marginal costs of precaution. The Economic Analysis of Law (AED in Portuguese) and a brief review of the theories of civil liability (subjective and objective) prescribed by the Consumer Protection Code (CDC in Portuguese)[9] and by the Civil Code (CC) were used to reach this result.

Finally, regarding the AI governance axis, proposals are presented for internal governance structures and measures, determination of the level of human involvement in AI-based decision-making, operations management, interaction and communication of the interested parties, both in public and private institutions, based on the Artificial Intelligence Governance Framework from Singapore. The methodology used for this was monographic research. In conclusion, the EBIA is a mechanism capable of inserting Brazil into the AI regulatory race but requires strong and effective public policies aimed at its implementation.

2 Ranking of Brazil in terms of AI in the international scenario

The classification of Brazil according to international indexers can bring considerations about the positive and negative impacts of the new economy and the information society with AI. In other words, this analysis expands the capacity to postulate public policies and outline which sectors, people, institutions, or companies have the greatest productive capacity or produce products and services linked to AI. These international experiences demonstrate that such processes may or may not be raising productivity at the local level with global impact from another perspective.

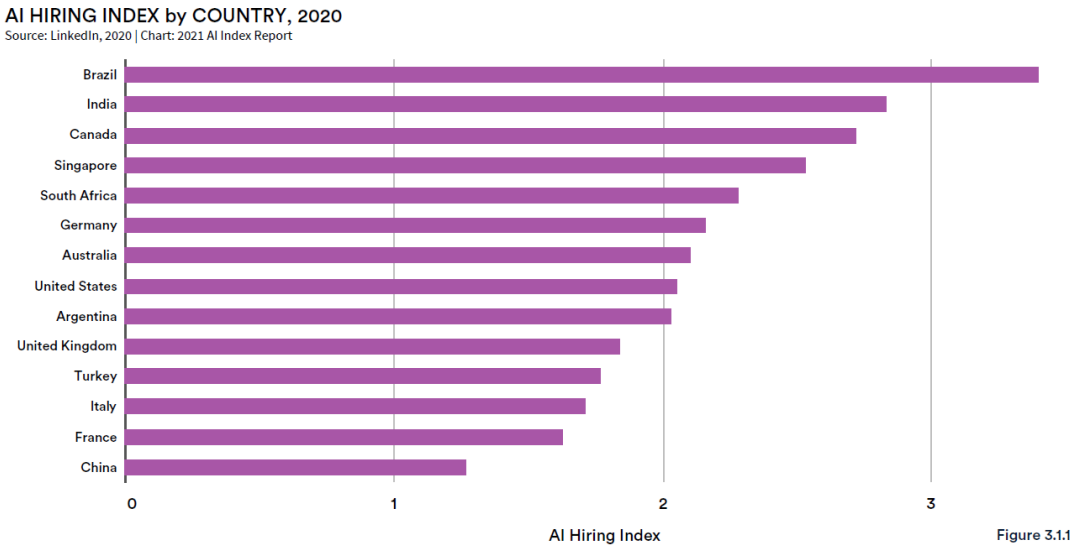
Because innovation is an essential and fundamental point for the country's economic development and is linked to the development of AI since it is a technique dependent on research, infrastructure, and development, the first step is to resort to the Global Innovation Index for the year 2021[10] (WIPO, 2021). Brazil occupies the 57th position of the global ranking in the general classification, the best since 2012, representing a higher-than-expected advance compared to 2019, when it occupied the 66th position, and 2020, in the 62nd position. Only Chile (53°), Mexico (55°), and Costa Rica (56°) from the Latin American and Caribbean regions are ahead of Brazil (WIPO, 2021).

These are the main points to be developed for the coming years concerning institutional apparatus (78°), infrastructure (69°), creative products (61°), and market sophistication (75°). The human capital resources destined for research (48°), business sophistication (34°), and knowledge and technology products (51°) should also be improved but have good positions in the ranking (WIPO, 2021).

[9] For more, see: (Divine, 2021).
[10] This index has 132 countries evaluated, among which are Switzerland, Sweden, Japan, Greece, Romania, Egypt, Bangladesh, Mali, Togo, etc.

In favorable terms, according to Artificial Intelligence Index Report developed by Stanford University (UNITED STATES, 2021), in 2020, Brazil was among the countries with the highest employability rate in the field of AI, along with India, Canada, Singapore, and South Africa (Figure 2).

Figure 2 – Employability index in the AI sector

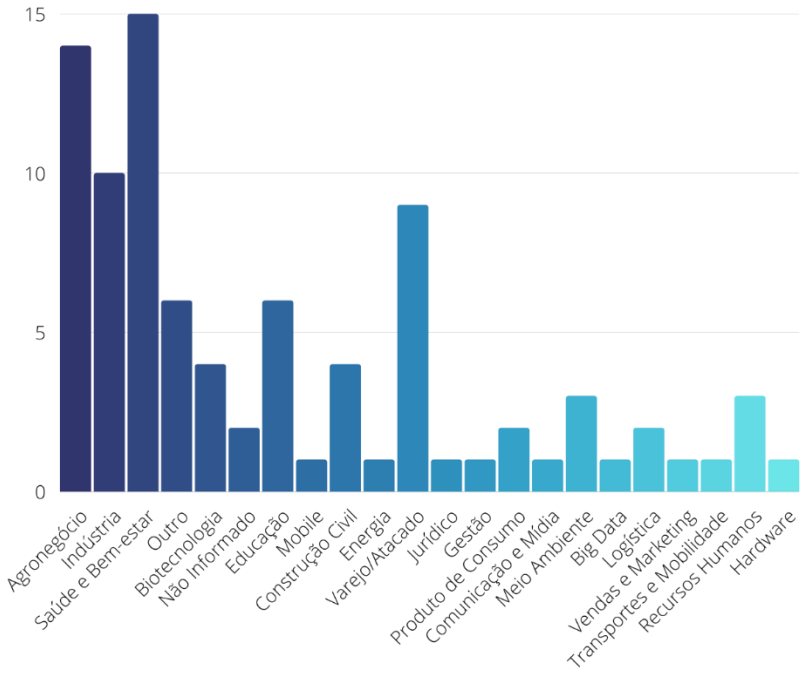


Source: United States (2021).

But the road is still long. When analyzing the data brought by EBIA, Brazil (2021) had approximately twelve thousand startups in 2020, with only 26 related to the AI sector. Updated data are presented for 2021 to update the data using the same government programs to promote innovation and entrepreneurship at the level of startups brought by EBIA – Conecta Startup (2021) and Startup Brasil program (BRASIL, 2021).

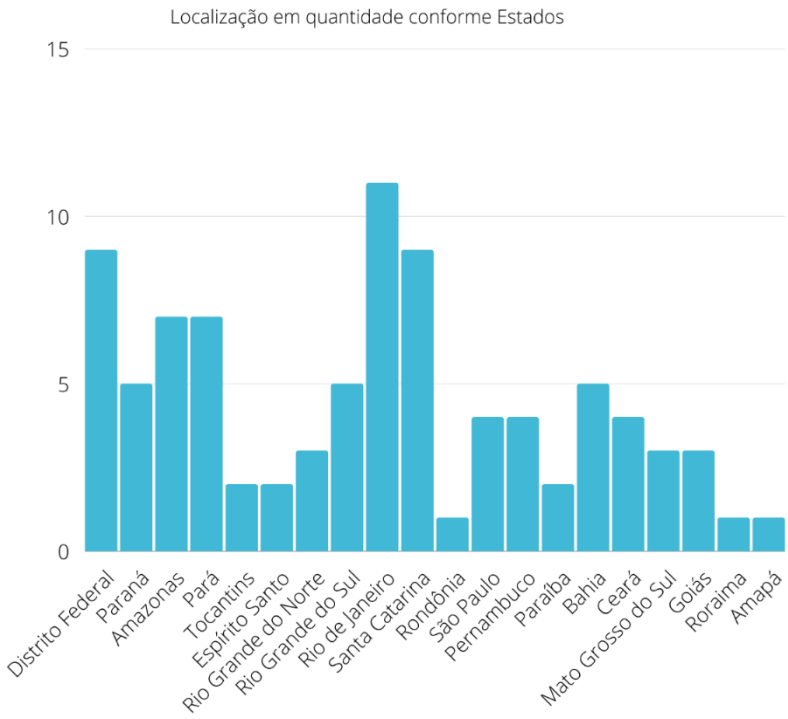
According to Conecta Startup, there were 94 registered startups. Figures 3 and 4 show that they predominantly occupy the agribusiness (14), industrial (10), and health and welfare (15) sectors, with most (not from these sectors) located in the Federal District (9), Rio de Janeiro (11), Santa Catarina (9), Amazonas (7), and Paraná (7).

Figure 3 – Areas of activity (in quantity) of *startups* according to the Conecta Startup program



Source: Prepared by the author based on the data provided in Conecta Startup (2021).

Figure 4– Areas of activity (in quantity) of *startups* according to the Conecta Startup program



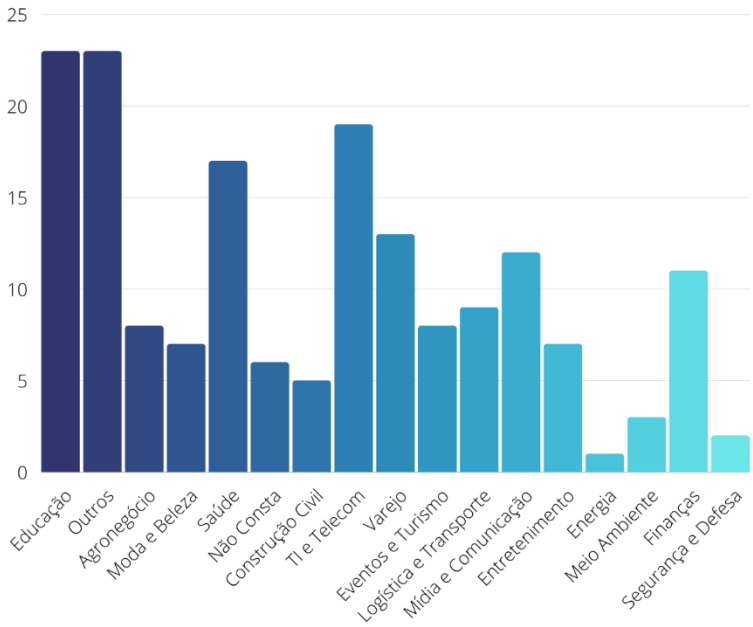
Source: Prepared by the author based on the data provided in Conecta Startup (2021).

The figures demonstrate a series of limitations and points that need improvement, among which is the greater effort in developing public policies to promote activities related to AI given a single startup connected to technology (hardware).[11]

The result obtained in the previous platform curiously emerges when analyzing the Startup Brasil Program. The data collected demonstrate the existence of 174 startups registered on their platform. They are predominantly in the education (23), health (17), IT and telecommunications (19), retail (13), finance (11), and media and communication (12) sectors, with most concentrated in São Paulo (54), Rio de Janeiro (16), Pernambuco (12), and Minas Gerais (10). The information is available in Figures 5 and 6.

Even though the area of IT and telecommunications is considerably larger regarding the previous program (19 to 1), the study is also limited by the impossibility of verifying whether the startups mentioned above work effectively in the AI sector. However, the partial results consider that, due to the range of the Brazilian territory and population, the numbers are tiny compared to the claims and results in the long term.

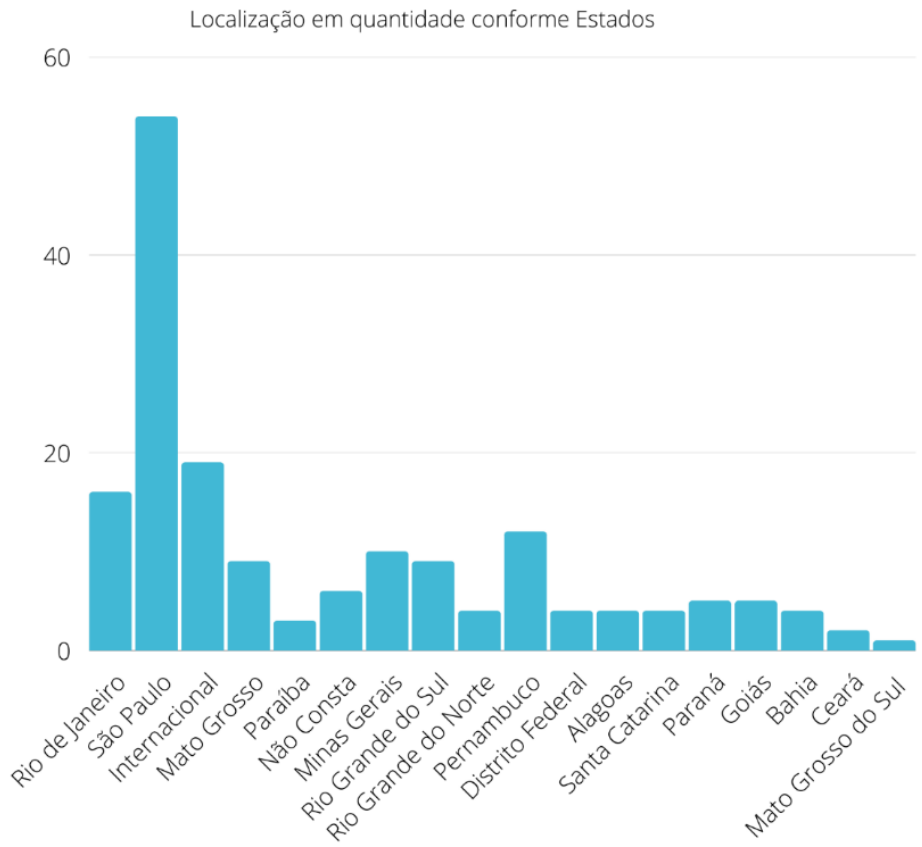
Figure 5 – Areas of activity (in quantity) of *startups* according to the Startup Brazil Program



Source: Prepared by the author based on the data provided in Startup Brasil (BRASIL, 2021).

[11] At this point, the study is limited by not expressly defining whether it operates in the field of AI.

Figure 6 – Areas of activity (in quantity) of *startups* according to the Startup Brazil Program

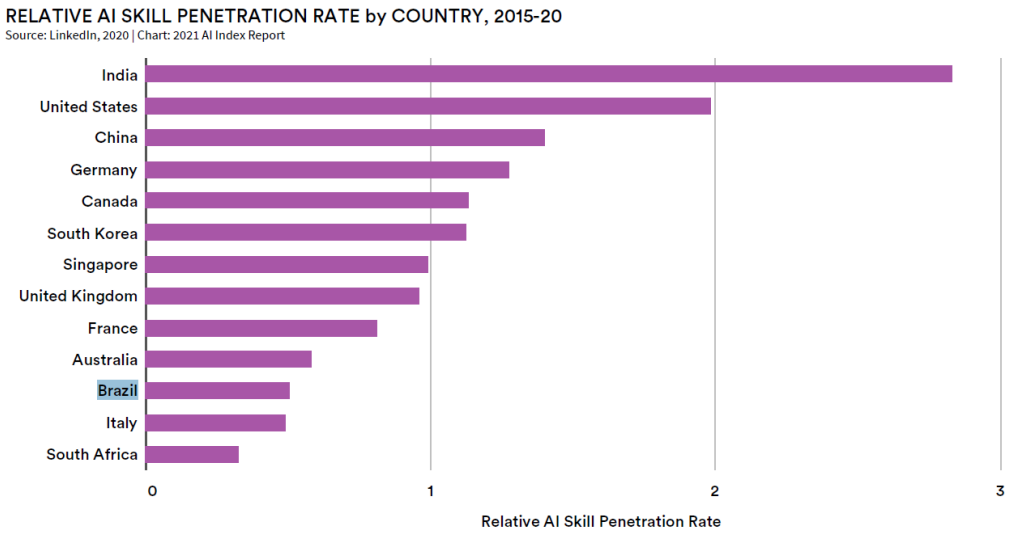


Source: Prepared by the author based on the data provided in Startup Brasil (BRASIL, 2021).

One of the essential results obtained through data collection is the observation of the diffusion of startups throughout Brazil, which demonstrates considerable entrepreneurial potential in Brazilian culture. However, the main problems faced by Brazilian startups are the scarcity of skilled labor, the high tax burden, and bureaucracy (BRASIL, 2021). The basic principle for its operation is the development of innovative and disruptive solutions to classic or new problems to break standards concerning companies in the same segment. This characteristic expressly requires an inventive and creative relationship in which Brazil is still precarious despite its continuous development.

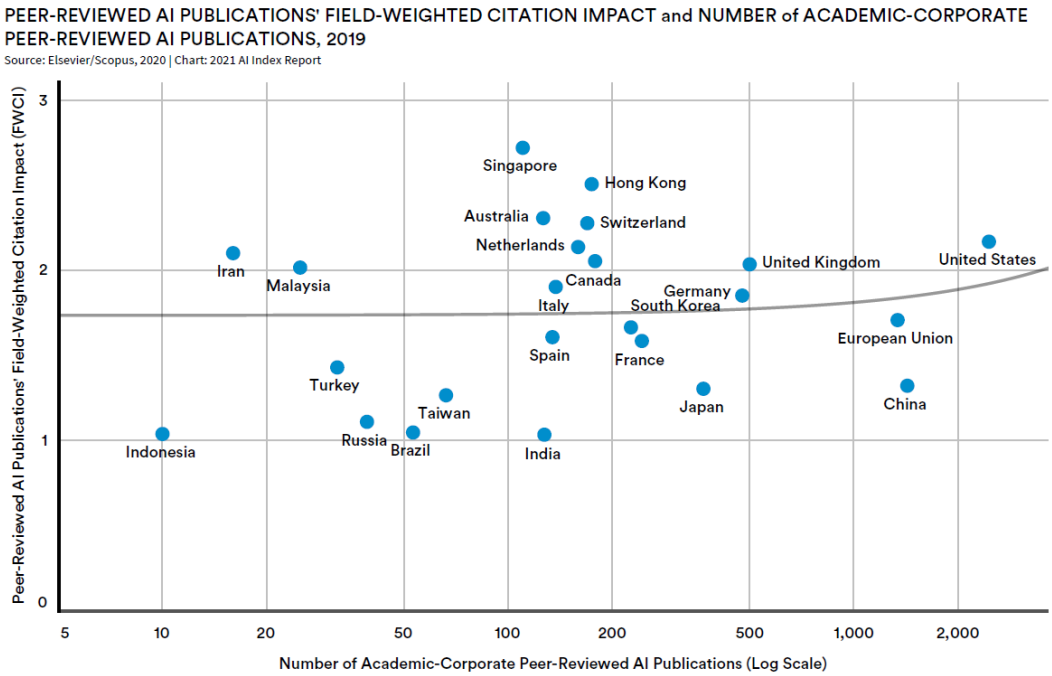
According to the Artificial Intelligence Index Report developed by Stanford University (UNITED STATES, 2021), the AI skills of Brazilians have low adherence, development, penetration, and prominence. However, they are still in the 50th position globally (see Figure 7), with Universidade Federal do Rio Grande do Sul being a reference in production related to AI ethics, machine learning, and conferences on robotics.

Figure 7 – AI skill index



Source: United States (2021).

Figure 8 – Publications made by the double-blind review system on AI



Source: United States (2021).

How EBIA based itself on the five principles defined by the OECD for the responsible management of AI systems, which are: 1) inclusive growth, sustainable development, and well-being; 2) human-centered values and equity; 3) transparency and explainability; 4) robustness, security, and protection; and 5) accountability. Each axis can and should be developed through public policies appropriate to its pretensions. Finally, the results above demonstrate that, no

matter how promising or how many expectations there are, the Brazilian sector requires improvements to achieve higher rankings and, primarily, better population satisfaction.

3 Thematic axis 1: Legislation, regulation, and ethical use

In this thematic area, EBIA's main concerns are related to the 1) protection of personal data; 2) prevention of discrimination and algorithmic bias; 3) balance aimed at preserving adequate structures to encourage the development of AI; 4) creation of legal parameters aimed at legal certainty regarding the liability of the different actors in its production chain.

The LGPD may initially address concerns 2 and 3. When the data processing is done in its automated form, the holder may request the controller or the operator to explain how it was carried out and its purposes.[12] The defense for explainability in AI systems comes from the assumption that its decisions may not necessarily be considered objective, fair, or impartial (BECKER; FERRARI, 2018; BURELLI, 2016; DIAKOPOULOS, 2013; PASQUALE, 2015). Thus, creating parameters to make explainability and transparency effective are basic mechanisms that must be fundamentally linked to avoiding bias in automated decisions and discrimination in data collection and processing (FJELD, 2020).

Its realization is based on the assumption of the exposition of the determining fundamental logic clearly and objectively, responsible for elaborating the decision. It is the presentation of a description that is understandable to the holder to delimit how the operator or controller used AI techniques to obtain the result (DOSHI-VELEZ; KORTS, 2017). In other words, what is intended is the verification of the parameters and criteria used in the automated processing and whether they have been correctly used to avoid errors or discrimination of the decision. One should verify in the algorithmic constitution (by design) 1) what are the main reasons that led to that decision; 2) whether the change, substitution, alteration, or modification of the factors would alter the decision; and 3) whether there are similar cases with different decisions, etc. (DOSHI-VELEZ; KORTS, 2017).

One of the forms to realize explainability in automated decision-making is by establishing the obligation for those responsible for development to leave the code of their software open (DIAKOPOULOS, 2013). If this is not possible, the algorithm should be specifically audited through cooperation between companies and the government, aiming to compare the result with the expected behavior in the constitution of AI (SANDVIG, 2014).

Therefore, the relationship between transparency and accountability permeates the need to adopt measures aimed at understanding the processes associated with automated decision-making to enable the identification of biases involved in the decision-making process. It turns

[12] Art. 20. The data holder has the right to request a review of decisions made solely based on automated processing of personal data that affect their interests, including decisions aimed at defining their personal, professional, consumer, and credit profile or aspects of their personality (General Data Protection Law - LGPD in Portuguese).

out that this practice can be effective and implemented based on the idea that AI systems should be human-centered (human-centric AI), which could facilitate its audit and the creation of more reliable systems (trustworthy AI).

One of the biggest challenges aimed at implementing this objective is the lack of technical knowledge by the general population to understand, interpret, and read the code responsible for executing the AI. A palliative solution, which should not be considered definitive, is the establishment of the duty to inform suppliers and those responsible about how, where, and when the decisions being made, if automated, are produced (OSOBA; WELSER, 2017).

EBIA's commitment to achieving this objective aligns its regulatory governance framework to creating public policies related to the topic and presenting initiatives considered relevant for its implementation. The first is the Brazilian Strategy for Digital Transformation (e-Digital) (BRASIL, 2018a), which seeks to coordinate the various government initiatives related to the theme around a singular vision, synergetic and coherent, to support the digitization of production processes and training for the digital environment, promoting, together, the generation of value and economic growth (BRASIL, 2018a). In total, 100 actions are planned, ranging from 1) connecting 22,000 urban and rural public schools with high-speed broadband access, in terrestrial or satellite networks, within the framework of the Connected Education Program; and 2) expanding the engagement of research and development centers in multilateral forums for defining international standards and radio frequency bands to be established for the fifth generation of mobile telephony (5G) to optimizing policies aimed at the sector to expand (more than proportionally) private investment in R&D improve the competitiveness of the Brazilian economy, generate more jobs with added value, and promote greater social development (BRASIL, 2018a).

The EBIA also indicates the Startup Brasil Program, aimed at supporting Brazilian startups and international organizations to develop software, hardware, and IT services or use these technologies to innovate. However, as previously visualized, few companies are linked to the program, and those destined to develop the intended area are still smaller.

In this context, it is believed that creating a National Center for Artificial Intelligence (CNAI in Portuguese) is appropriate to assist the connection and contact between society and government institutions through AI experts. In addition, the CNAI will be able to act specifically in projects aimed at creating, using, and improving AI in multiple sectors of the economy. This practice would facilitate collaboration between the industrial and research sectors, aiming to attract more significant investments due to qualified labor.

Another challenge in the regulatory sector related to AI is the ethical and civil liability aspects. The EBIA does not postulate how we should resolve this issue, but specific considerations will be presented below.

3.1 Contributions and proposals in AI civil liability

I stress that the possibility of an AI being or not a subject of law is not addressed, but rather the ethical and legal responsibility arising from its acts. Attention, in this scope, will be directed to a disruptive landscape in which AI is incidentally in praxis and can come to cause damage to civil society. Even when we try to demonstrate the requirements 1) agent; 2) damage; 3) fault; and 4) causation, the former is unverifiable. Thus, I understand that an AI cannot be personally responsible for the damage caused to third parties, at least at present. The intention is to verify how the theories of civil liability may or may not be enough to treat the current problem.

Microsoft had inserted a chatbot called Tay to interact autonomously with users of this social network. Its development was pretentious: through ML, Tay should be able to compile information in its database to create understandable discourses (PEREZ, 2016). However, the company disabled the AI after reports that Tay would send racist and misogynist tweets (VINCENT, 2016). In this light, shortly after that, the Microsoft-like Facebook project was shut down as it realized that its two AIs (Alice and Bob) were interacting with each other in an unintelligible way.[13]

Although they were controllable situations and easy to verify and remedy, the AIs had an offensive potential that could develop in the chain of two large social networks. This development could reach extremist groups and serve as a basis for discriminating against other networked groups.

When we analyze the cases of applicability of liability for the fact of the product in the CDC, we usually link it to personal and property damage of a physical nature. In some cases, even the loss or deterioration of the product due to its malfunction. This product is usually physical, although there are possibilities for digital products. However, we go beyond physical damage when considering AI in this context. An AI, for example, can cause damage to the privacy of the holder of that product. Amazon, for example, collects intimate details about its users through Alexa, its AI.[14]

Another example is electronic and intelligent locks that can remove the autonomy of

[13] Part of the dialogue can be expressed: “A few days later, some coverage picked up on the fact that in a few cases the exchanges had become - at first glance - nonsensical: Bob: "I can can I I everything else" Alice: "Balls have zero to me to me to me to me to me to me to me to me to” (BARANIUK, 2017).

[14] "Amazon collects data about consumers through its Alexa voice assistant, purchases on its marketplace, Kindle e-readers, and its music platform. The company gathers a vast array of information about its U.S. customers and began making that data available to everyone upon request early last year after trying and failing to defeat a 2018 California measure that required such disclosures. [...] A reporter's dossier revealed that Amazon collected more than 90,000 Alexa recordings from family members between December 2017 and June 2021 - an average of about 70 per day. The recordings included details such as the children's names and their favorite songs. Amazon captured children asking how they could convince their parents to let them play and receiving detailed instructions from Alexa on how to convince their parents to buy video games. Some recordings involved conversations between family members using Alexa devices to communicate in different parts of the house. Several recordings captured children aged seven to 12 asking Alexa questions about terms such as "pansexual." The reporter did not realize that Amazon was storing the recordings." (O GLOBO, 2021, our translation).

their holders if third parties use them. In 2017, a Hotel in Australia was attacked by *hackers* who managed to break into the company's security system, specifically the electronic lock system. A sum of AUD 1,800 was required from the guests to re-enter their room and recover their belongings (BILEFSKY, 2017).

It turns out that liability in the CDC has some specificities, among them objectivity and risk-based reasoning. We begin with the second. The relationship between risk and activity is always probabilistic. In other words, according to this author's understanding, the central pillar of the risk theory in the CDC is the possibility of predicting or not the existence of damages. This prediction is not easy to establish within the framework of AI. We have seen Bob and Alice's relationship as entirely outside the pre-established standards. Deep Learning (DL) and Machine Learning (ML) are necessarily made not to predict an AI's behaviors. After all, if the intention is to replicate the behavior of the human being, nothing is more appropriate and inspiring than to replicate its randomness.

In a way, predictability through ML systems is not denied. Some rules can be entered into the AI system so that they act as predetermined. However, these rules have semantic and syntactic boundaries. Moreover, the greater the number of rules, the greater the complexity and, accordingly, the higher the costs of operating such AI. Filling the content of an AI with all possible rules in computational linguistic and probabilistic content seems to go beyond the pretensions of ML.

It is stressed that not all AI acts are necessarily unpredictable, such as an autonomous vehicle that stops at a pedestrian crossing for a passerby. The issue is that the risk of causing harm is not merely technical but also a normative risk, escaping the scope of the CDC.

Thus, it is believed that the risk of AI development activity far transcends the risks protected by the CDC. But in what form does this occur? First, we must review some concepts. The liability for the risk of the product tries to solve some problems, namely, 1) defects arising from the design, manufacture, construction, assembly, formulas, handling, presentation, or packaging of its products; and 2) insufficient or inadequate information about its use and risks. It concerns the strict liability for damages, even if unintentional, to the consumer, whether they derive from the defective product or the lack of information regarding its use and risk.

If an AI did not work as it initially should, it would be enough to verify in its constitution, design, or manufacture, for example, that it is a product defect and, consequently, would attract liability to the manufacturer, the producer, the builder, national or foreign, and the importer.[15] The defect is the key to realizing the liability of the above subjects before the consumer.

[15] According to art. 12, paragraph 1 of the CDC, "The product is defective when it does not offer the safety that is legitimately expected of it, considering the relevant circumstances, among which: I - its presentation; II - the use and risks that are reasonably expected of it; III - the time in which it was put into circulation."

The CDC established a few approaches to verify defects in the product: presentation, use, risks, and time it was put into circulation. This means that the manufacturer must prove the absence of this defect in at least three legally established modalities. From the economic aspect, the person responsible must adopt measures that reduce the risk imposed on the development of the product through precautionary and damage prevention techniques that could be the key to a wrongful act. It concerns the precautionary costs defended by the AED, the non-observance of which attracts the strict liability to the agent.

It turns out that AI may not have any defects and, even so, cause harm to the consumer. This is where the challenge begins. One of the excludents of strict liability of art.12, paragraph 3 of the CDC is the proof that the product was placed on the market and the defect does not exist.

In addition, there is a **safety expectation about the product** (art. 12, paragraph 1 of the CDC), which becomes almost impossible to verify in practice due to technological novelty. Within this theme, how does one draw sufficient and adequate information about the possible risks of the product (art. 12, head provision, final part, CDC) considering the main problem of AI as its unpredictability and, consequently, the blurring of risks in probabilistic terms (LEMLEY; CASEY, 2019)? In other words, if unpredictability is part of the product, how does one insert and accept this unpredictability as part of the risks of the activity, being that it is unknown? It would be difficult and economically unfeasible to include all the operational risks of such AI in the database. Otherwise, if the manufacturer for the unpredictable risks were punished, we would turn them into predictable and recognize their responsibility in all possible cases. In other words, they would never be able to exclude the wrongfulness contained in art. 12, paragraph 3, II of the CDC because the technology would be defective. Reasoning, on the contrary, can be verifiable and intelligible. Rachum-Twaig prescribes that

> since AI-related risks are unpredictable in nature and, therefore, cannot be covered by the design defect or the duty of warning and instruction doctrine, there may be cases of damages outside the scope of the product liability doctrine; however, these cases can be offset by other forms of tort liability (RACHUM-TWAIG, 2020, p. 1141, our translation).

Another problem arising from the risk analysis not only in the CDC is strict liability. Unlike fault liability based on guilt, strict liability is essentially verifiable in risk.[16] The general problem of the absence of predictability in AI behavior is crucial for an adequate normative interpretation since its incidence is indispensable to adapting and reallocating risks under the prism of strict liability. In addition, there is a latent informational asymmetry between manufacturers, developers, programmers, and everyone involved in the chain. To a large extent, developers know and have more information about the risks of AI than manufacturers. These also have a considerably higher level of information regarding the AI than the final recipients.

[16] "fault liability makes wrongful agency the fundamental basis of responsibility for harm accidentally done; strict liability makes agency itself the fundamental basis of responsibility" (KEATING, 2001, p. 1285).

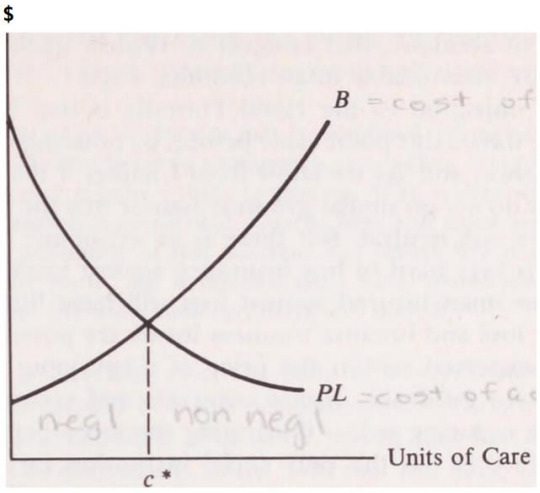
Rachum-Twaig (2020, p. 1163) brings examples of a robot acting in medicine that has more information on the procedure's risks than the patient to be operated on. Because the primary characteristic of AI-based robots is that they can act unpredictably or inexplicably toward humans, often none of the stakeholders will be better placed to assess the risks involved in their operation, and the problem of imperfect information will apply equally to all stakeholders. In such cases, the guiding principles of applicability of strict liability cannot be applied.

The reduction of costs in the production chain will be directed exclusively to technology designers and developers, who will have to obtain any and all types of information related to the risk to mitigate and eradicate it as far as possible. Programming in this way becomes unfeasible under the prism of the relationship between costs and benefits. Unpredictability is intrinsic in electronic programming, and its risks must be understood beyond those established in the CDC and strict liability for not being predictable.

Therefore, I understand that adopting the CDC in these cases would only hinder and transform the technological environment into a normative environment endowed with rules and legal precepts without practical applicability and only with negative reflections. Theoretically, we could conclude that the regime of fault liability would be ideal since strict liability should be set aside. From the perspective of AED, Hand's formula can be used as a mechanism aimed at solving this problem. However, we will see that it is not so easy.

According to Posner, Hand's formula imposes marginal precautionary costs (B) on the parties to the legal relationship to avoid possible expected damages (PL) arising from a multiplication between the probability that this damage would occur (c) and the damages (d) (POSNER, 2012).[17] It is represented as follows:

Figure 9 – Prevention Costs



Source: Posner (2012).

[17] For more on AED see (SHAVELL, 2009)

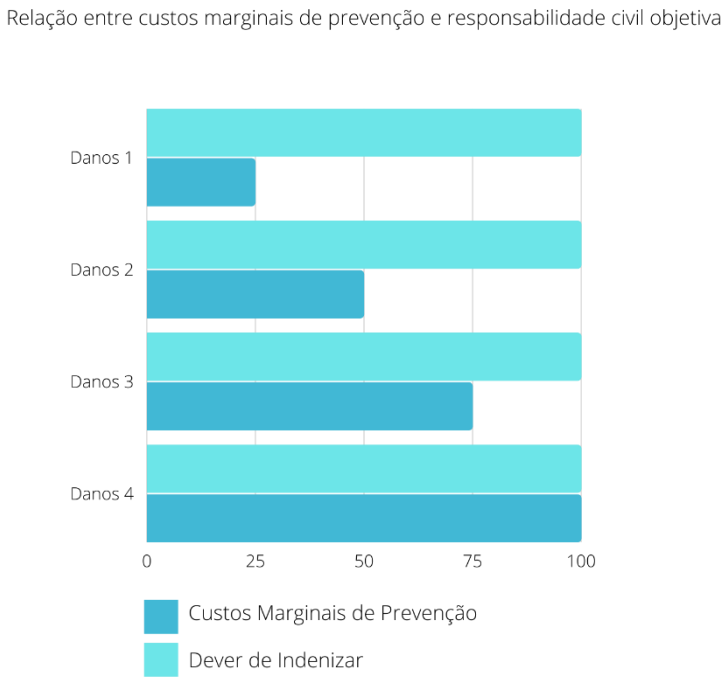
The graph above shows that, for Hand, the conduct will be culpable when the investment in precaution (B) is lower than the possible expected damages (PL). Thus, B < PL x c (POSNER, 2012, p.148). The decreasing curve represented by PL represents a marginal change in the expected damage costs as a function of the precaution adopted. That said, it turns out that adopting prevention and precautionary practices decreases accident prevention. On the other hand, the curve marked by B exemplifies the marginal cost of care. It rises to the extent that precautionary practices are scarce and, therefore, rises as more products and services are offered in the consumer market. The intersection between the curves consists of the ideal duty of care, where there is a separation between negligence and prudence/precaution (POSNER, 2012).

When this analysis is applied from the perspective of strict liability, if it is determined that the reparation of the damages must be made unconditionally and invariably by its developer, producer, manufacturer, or responsible for the technology, without fault analysis, the interests for the implementation of new practices aimed at preventing the occurrence of new damages would be denied and absent.

Therefore, it is a disincentive to adopt marginal costs of prevention. If the company followed all the precepts recommended by the legislation and adopted the practices of governance and precaution and, even so, was made liable for any and all damage that occurred in its contractual area, the costs of these actions would be dispensable, given that the liability would occur with or without them. Thus, by adopting strict liability, the marginal costs of precaution (B) to avoid possible expected damages (PL) are indifferent to the damages themselves (d).[18] They are expressed as follows:

[18] For a better understanding of AED, especially in Posner, we recommend (GAROUPA; PORTO; FRANCO, 2019).

Figure 10 – Marginal Costs x Strict Liability



Source: Prepared by the author.

It should be noted that, regardless of the actions taken, the indemnity criterion always permeates the totality of the damages incurred in strict liability. Thus, if this (strict) modality is adopted, it would make no sense for the developer or company to invest and adopt transaction costs aimed at prevention since they would be obliged to compensate the injured holders. Thus, this is an effective loss ratio and an unnecessary expenditure in business policy.

Under the economic and legal context, if Hand's formula is correct, adopting civil liability in its fault modality seems to be the most appropriate since it would allow the costs destined to the precaution not to be reduced to zero due to unconditional liability.

Glaubitz and Raymond (2021) recognize the possibility of Hand's formula in solving the problem of imputing liability for the acts of AI and insert the third variable in the probabilistic calculation: the duty of care.

Glaubitz and Raymond formula (2021) for AI liability verification:

$$B - \frac{k}{\min\{A\}} < P \cdot L$$

In the formula above, *B* is the duty and care to have the automated task completed

manually by a human; *k* is a constant that determines the influence of *A*; *A* is the minimum cost of an alternative algorithmic solution (requalification of the existing algorithm, creation of a new algorithm, or use of an existing algorithm that achieves the same legitimate purpose with a lower impact on protected classes); *P* is the likelihood of pre-existing duty of care, determined by a disparate impact assessment that assesses the extent of any disproportionate adverse impact on a protected class (identified through the algorithm audit); and *L* is the loss or damage arising from the algorithmic bias experienced by a protected class (GLAUBITZ; RAYMONG, 2021, p. 31).

In practical terms, *A* would be like a civil penalty. If the marginal cost of prevention is high, the perpetrator would be less penalized. If the cost incurred in the duty of care is low, compensation should be valued according to a fault. According to Glaubitz and Raymond (2021), liability will vary according to the degrees of automation of the product and its supervision. The more automated, the greater duty of care is required to mitigate the offensive potential to society. This duty of care would be complementary to the cost of prevention *P* previously existent on Hand.

However, the risk is an indispensable factor to be verified in the formulas of both Hand and Glaubitz and Raymond (2021). Risk is, in short, an element that must be objectively measurable to reduce it. This risk is complex – and in some cases impossible - to outline in its entirety for AI programming. Unable to verify *ex ante* actions to predict and avoid them *ex post*. Impact assessments are impaired as an appropriate instrument for applying the formula described by Hand, and consequently, we fall into an activity full of risks and uncertainties.

Therefore, in terms of civil liability, the fault modality seems more appropriate for implementing the EBIA precepts regarding maintaining innovation and stimulating its incentive. For this reason, public policies aimed at financing research projects that seek to apply non-discriminatory solutions based on equity/non-discrimination (fairness), responsibility/accountability, and transparency are the matrices for stimulating partnerships with companies researching commercial and social solutions of these technologies.

In other words, the State shall establish standards and technical requirements for promoting responsible AI. At the same time, that meets the forms of liability previously defined in legal standards. If necessary, mapping legal and regulatory barriers will be essential to identify and update the promotion of legal certainty. This practice will only be possible when more in-depth studies are carried out.

Therefore, as established by the EBIA, creating data quality control policies for the effectiveness of ML and DL under human interventional parameters can help find effective results with low risk to society and the individual. And again, creating a specific center for AI, with experts in the field, is essential.

We find ourselves in the governance sector when analyzing how the liability framework

can be implemented. At this point, the fundamental aspect is the creation of management practices for monitoring and supervising AI systems.

4 Thematic axis 2: AI governance

The EBIA, at this point, postulates two basic and guiding principles of the governance structure: transparency and accountability (here, brought in terms of government and understood as liability and accountability). These parameters require the person responsible for the application of AI to establish structures designed to ensure that its implementation can be analyzed by the precautionary principle, identifying high-risk applications that can significantly impact society in a given application context, such as health or monitoring of public space.

According to the EBIA, regulatory intervention regarding governance must be balanced between the degree of risk related to the specific application of AI and the possible limitations that may restrict its uses. To this end, the EBIA proposes the preparation of periodic data protection impact reports (RIPDs in Portuguese), which can be identified in each sector of AI activity, such as Security Impact Report (RIS in Portuguese); Environmental Impact Report (RIA in Portuguese); or Human Rights Impact Report (RIDH in Portuguese).

In the legal field, Decree nº 8.777/2016, responsible for the Brazilian Open Data Portal, and the Brazilian Public Software Portal, governed by Ordinance nº 46/2016, are already measures and policies aimed at project transparency. However, the governance system must begin from an assumption of structural collaboration between the public and private sectors to develop risk management standards associated with using AI.

In other words, there are four key steps to make an AI governance model effective when using Artificial Intelligence Governance from Singapore:

> a. Internal governance structures and measures designed to adapt existing structures and measures or establish new ones to incorporate values, risks, and liabilities related to algorithmic decision-making.
> b. Determine the level of human involvement in AI-based decision-making: a methodology to help organizations establish their appetite for risk for using AI, i.e., determine acceptable risks and identify an appropriate level of human involvement in AI-based decision-making.
> c. Operations management: issues to consider when developing, selecting, and maintaining AI models including data management.
> d. Interaction and communication of the interested parties: Strategies for communicating with an organization's interested parties and managing relationships with them (SINGAPORE, 2020, our translation).

It is a flexible model which public or private organizations can change according to their needs and cultural aspects. The important thing, in this case, is that the guidelines can assist and help organizations understand how to implement each of the practices described above.

4.1 Internal governance structures and measures

Internal governance structures and measures aim to ensure robust supervision over the use of AI by an organization, whether public or private. Existing internal governance structures can be adapted, and new structures can be implemented if necessary. For example, risks associated with using AI can be managed within the entrepreneurial risk management framework. At the same time, ethical considerations can be introduced as corporate values and managed through ethical review boards or similar structures (SINGAPORE, 2020).

Some factors are relevant to its creation and effectiveness:

1) Clear roles and liabilities for the ethical use of AI since it is developed in stages and by activities. There may be an individual allocation of responsibility to its developer in each of them. Thus, departments with an internal governance structure can know their roles and be fully aware of their functions and liabilities. Among the practices adopted are a) the maintenance, monitoring, documentation, and review of AI models developed; b) the review of communication channels and interaction with users to provide effective feedback; and c) ensuring that employees/developers are technically able and properly trained to handle AI systems (SINGAPORE, 2020).

2) Risk management and internal controls are an option to identify, correct, and address internal risks when developing AI. Reasonable efforts may be made to ensure that the datasets used for AI model training are fit for the purpose, assess and manage inaccuracy or bias risks, and review exceptions identified during model training. Virtually no data set is entirely unbiased. Organizations should strive to understand how datasets can be biased and address this in their security measures and deployment strategies (SINGAPORE, 2020).

One of the examples capable of illustrating the implementation of internal governance measures and structures is that adopted by the Mastercard company. To ensure robust oversight of Mastercard's use of AI, the company has established a Governance Board to review and approve the implementation of AI applications determined to be high-risk. The Governance Board is chaired by the Executive Vice President of the Center for Excellence in Artificial Intelligence. Their members include the Chief Data Officer, the Chief Privacy Officer, the Chief Information Security Officer, data scientists, and commercial team representatives (SINGAPORE, 2020).

4.2 Determination of the level of human involvement in AI-based decision-making

EBIA proposes a Human-Centric AI approach. In simple terms, this approach places AI as a tool where human agents decide when and how to use it. According to Muller, "we require a HIC approach to AI, where machines remain machines and people maintain control over these machines at all times" (UNIÃO EUROPEIA, 2018). In its conception, "human agents can and should have control of whether, when, and how AI is used in everyday life, as well as what

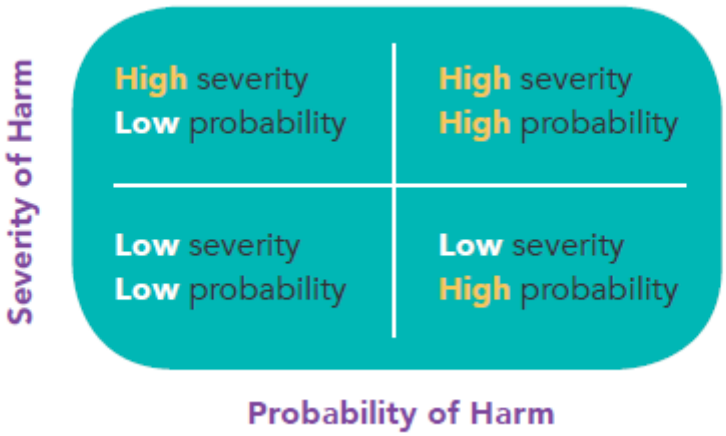
tasks we transfer to AI, how transparent it is, and respect for ethical aspects" (UNIÃO EUROPEIA, 2018). Therefore, it is assumed that the monitoring and supervision of the AI system from its conception (privacy by design, security by design, human rights by design, ethics by design).

To achieve these objectives, companies and the government must decide on their commercial objectives before making AI available on the market and verify the risks in their use and decision-making. Different cultural factors and normative or value systems must be considered for multinationals. Additionally, some risks for certain subjects may manifest only when applied in a specific group (such as the automated offer of products that can cause behavioral change and, consequently, compulsive purchases).

Moreover, as the identification of commercial objectives and risks and the determination of human action in automated decision-making are interactive and continuous processes, it is expected that organizations continue to identify and constantly review them to improve solutions in their technologies, therefore, mitigating risks and maintaining an effective response to previously failed actions.

According to considerations extracted from the EBIA, it becomes interesting to think of automation not as the absence of human involvement in a given task but to include it selectively in the expectation that the result will be an effective process of the characteristics of intelligent automation (WANG, 2019).[19] This structural model begins from two major axes: 1) probability and 2) ability to cause harm to an individual (or organization) due to the decision obtained in the process.

Figure 11 – Damage x probability ratio



Source: Singapore (2020).

[19] For more (DIVINO; MAGALHAES, 2020).

The proposal of foreign authors is standardized and similar to that already existing in the general theory of government. In the public sector, in the same way, as in the private sector, the manager must use tools that can assist in decision-making. Tools are even more critical when it comes to environmental, health, economic, ergonomic, and other risks that must be managed and controlled in institutions in all government areas since the welfare of society is the essence of public service. It is essential to make decisions and use correct measures concerning public policies and programs and to adopt effective risk management strategies.

One of the primary regulations to assist the public manager in the exercise of their position and function as a person of the people is the Joint Normative Instruction CGU/MP nº 1 of 2016. This standard seeks to ensure that decision-makers at all levels of the body or entity have timely access to sufficient information regarding the risks to which the organization is exposed. For this, it must be understood that the posture of the chief executive disseminated in their social networks or "private" environment does not reflect only in their subjective sphere but also in that of those who adopt the same positioning externalized by them. For this reason, the behaviors practiced, even under the spectrum of freedom of expression, must consider the impact on society, especially in the critical area of health.

Risk management is based on the assumption of impact x probability. When an unfavorable position of vaccination is externalized, its impact on society and the likelihood of this happening should be verified. According to the Federal Accounting Court, it can be exemplified as follows:

Figure 12–- Risk assessment method

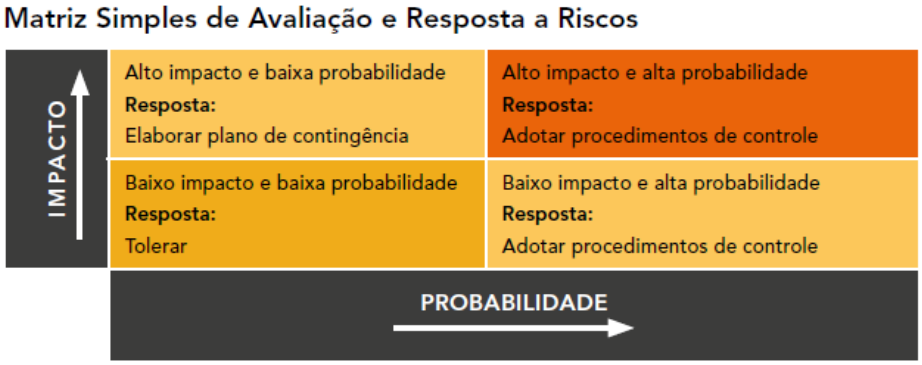


Figura 3: Matriz 2 x 2 de resposta a risco (INTOSAI GOV 9130, traduzido e adaptado)

Source: Brazil (2018b).

The greater the impact and likelihood, the greater the effort required to avoid it as much as possible. It is, therefore, a directly proportional relationship aimed at mitigating risks.

Figure 13 – Risk management model

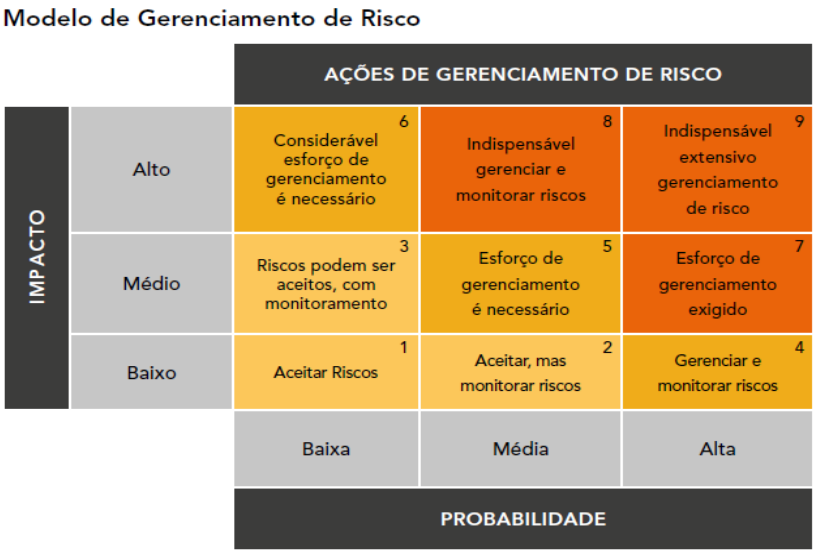


Figura 4: Matriz 3 x 3 de gerenciamento de risco (Secretaria do Tesouro do Canadá)

Source: Brazil (2018b).

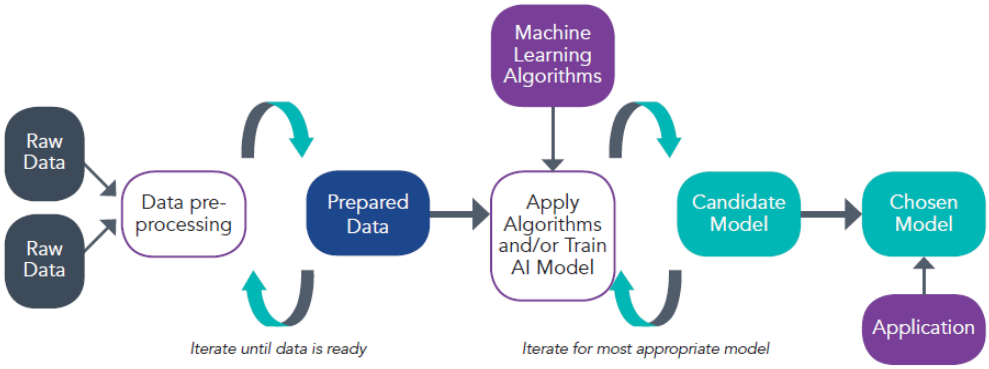
With these practices in place, there is considerable opportunity for improving the quality of AI systems at the government and private levels. Concerning public policies in the state sector, data protection impact reports can acquire new features and increase concrete results to promote AI in conjunction with its liability and ethics. This social and multisectoral dialogue is essential to leverage the practices of accountability related to AI in organizations.

4.3 Operations management

A series of processes and operations must be intrinsically linked to the organization of data, algorithmic, and of the very constitution of AI for the practices of the previous topic to be implemented. The EBIA, however, does not establish such criteria. However, the Artificial Intelligence Governance Framework of Singapore proposes the organization in three steps: 1) data preparation, 2) algorithm analysis and 3) choice of the appropriate model for the proposed.

Data collection is done in the first stage. It is formatted and treated to obtain more assertive conclusions. In this case, the accuracy and insights increase as the amount of data in the AI training base increases. These models are then trained in the database, and an algorithmic analysis is performed. One can include statistical analysis or machine learning models extracted from neural networks. The results are examined and entered into the most appropriate models. To this end, the analysis will be probabilistic and incorporated into applications to offer predictions, make decisions, and solve problems.

Figure 14 – Operational Management Model in AI



Source: Singapore (2020).

The whole process undergoes the linear understanding of all the data, from how, when, and where they arrived, how it was collected, treated, and transferred, and where, when, and where they will go. The key to a transparent procedure before the other bodies will be the existence of records that allow the organization, whether public or private, to guarantee the quality of the data according to its origin. In the meantime, practices that minimize discriminatory biases in decision-making will also be fully adopted. One of the most common biases to be addressed is a racial and stereotypical issue and the omission of certain people according to their ancestry in an economic or social group. For this reason, the relationship between different databases for training, testing, and validation of AI in public and private institutions is essential to ensure a periodic review and constantly updating the results obtained.

Singapore's regulatory proposal directs organizations to a risk-based algorithmic modular application represented in two stages. The first should identify the subset of features or functionalities that significantly impact the interested parties for whom such measures are relevant. Second, identify which of these measures will most effectively build trust with your interested parties. Some of these measures, such as explainability (or repeatability when using models that are not easily explained), robustness, and regular tuning, are sufficiently essential that they can, to varying degrees, be incorporated as part of the organization's AI implementation process. Other measures, such as reproducibility, traceability, and auditing, are more resource-intensive and may be relevant for specific characteristics or in specific scenarios (SINGAPORE, 2020).

In this same sense, explainability can be achieved when outlining how algorithms work or how the decision-making process incorporates prediction models. This correlates with understanding and trust in the systems being developed and implemented. However, as already

mentioned, there are still technological and social limitations regarding explainability. Where explainability cannot be achieved, organizations could consider documenting the repeatability of the results produced by the AI model. Repeatability refers to the ability to consistently perform an action or make a decision given the same scenario. Although repeatability (of the results) is not equivalent to explainability (of the algorithm), some degree of assurance of consistency in performance could provide AI users with a higher degree of confidence (SINGAPORE, 2020).

Other factors, such as robustness[20], continuous improvement[21], traceability[22], standardization and possibility of reproduction[23], and AI audit,[24] are indispensable for implementation and intended for linear and security procedural understanding.

Finally, public policies aimed at the development of AI are only possible through stakeholder interactions and communications, which, in this case, is civil society and public and private institutions.

4.4 Interaction and communication of the interested parties

Because the incorporation of AI in public power is based on the reference to strengthen human activities, the public power, in this case, has no interest in satisfying private interests, which would relegate the public interest. In other words, the implementation of AI in this sector must be done through the avenues of equity and social inclusion. To this end, government organizations must provide general information on which sectors, products, and AI services applications will be effectively used. At this point, developing policies and terms of use is

[20] "Robustness refers to the ability of a computer system to cope with errors during execution and erroneous input and is assessed by the degree to which a system or component can function correctly in the presence of invalid input or stressful environmental conditions. Ensuring that deployed models are sufficiently robust will contribute towards building trust in the AI system" (SINGAPORE, 2020).

[21] "Establishing an internal policy and process to perform regular model tuning is effective for ensuring that deployed models cater for changes to customer behaviour over time. This allows organisations to refresh models based on updated training datasets that incorporate new input data. Model tuning may also be necessary when commercial objectives, risks, or corporate values change" (SINGAPORE, 2020).

[22] "An AI model is considered to be traceable if (a) its decisions, and (b) the datasets and processes that yield the AI model's decision (including those of data gathering, data labelling and the algorithms used), are documented in an easily understandable way. The former refers to traceability of AI-augmented decisions, while the latter refers to traceability in model training. Traceability facilitates transparency and explainability and is also helpful for other reasons. First, the information might also be useful for troubleshooting, or for an investigation into how the model was functioning or why a particular prediction was made. Second, the traceability record (in the form of an audit log) can be a source of input data that can be used as a training dataset in the future" (SINGAPORE, 2020).

[23] "While repeatability refers to the internal repetition of results within one's organisation, reproducibility refers to the ability of an independent verification team to produce the same results using the same AI method based on the documentation made by the organisation. Reproducibility can influence the trustworthiness of the AI product and the organisation deploying the AI model. As implementing reproducibility entails the involvement of external parties, organisations can take a risk-based approach towards identifying the subset of AI-powered features in their products or services that requires external reproducibility testing" (SINGAPORE, 2020).

[24] "Auditability refers to the readiness of an AI system to undergo an assessment of its algorithms, data and design processes. The evaluation of the AI system by internal or external auditors (and the availability of evaluation reports) can contribute to the trustworthiness of the AI system as it demonstrates the responsibility of design and practices and the justifiability of outcomes. It should, however, be noted that auditability does not necessarily entail making information about business models or intellectual property related to the AI system publicly available" (SINGAPORE, 2020).

strongly recommended. They should contain explanations to help users communicate with AI and request information from the state entity responsible for its development and use. The policy may also contain its different functions and the individualized liability of each member responsible for elaboration.

Note that the detailed explanation can help communication between the public sector and considerably increase transparency in decision-making. Thus, both explainability and transparency are appropriate and indispensable mechanisms for interaction and communication between sectors, consequently increasing trust in these types of applications. Because the government sector is broad, the first step to outline this policy and the terms of service is to identify the target audience for the provision of services and then insert them in a context in which their claims and proposals are compatible with AI applications.

More general information can be directed to potential users, deciding whether or not to join the service run by the AI. If accepted, more specific information should be directed to demonstrate how the application works. Finally, a step-by-step should be indicated to your users if human intervention is required. If they request information on the operation of automated decision-making, they must be included in a feedback parameter.

In other words, creating feedback channels intended for user evaluation is indispensable for the organization, especially for the Data Protection Officer, to make their decisions and verify the adequacy of practices to legislation. According to the European Commission, all governments in constitutional democracies should be limited by law, including those that use AI systems to maintain or expand their democratic processes.

Note that the parameters established for implementing public policies, although individualized, constitute an inseparable procedural set to ensure the appropriate inclusion of the system in society. Security and transparency, as already evidenced, are fundamental instruments to maximize innovations between regulation and the economy.

Given the above, the EBIA has high expectations at the international level of implementation but requires high and continuous work at the level of public policies for its effectiveness.

5 Conclusion

The research problem proposed was how and what public policies can be adjusted for the effectiveness and implementation of EBIA regarding the legislation, regulation, and ethical use axes and governance of AI. The first section showed that, at the national level, EBIA has significant challenges to be met, among which is the greater incentive to research, technology, innovation, and development at the national level and for startups. In this case, the programs directed to the promotion are residually applied in the IT sectors. As visualized, one of the study's limitations with the data obtained is to trace the concrete scope of action. Also,

Information Technology is a vast sector. Thus, the results in practice can be even worse when concretely located in AI-related sectors.

Therefore, the primary contributions of this work lie in the finding that civil liability for illegal acts committed by AI or during its performance, execution, and implementation should be fault liability. Thus, it allows the mitigation of damage according to precautionary costs effectively allocated for configuring an ethical, solid, and robust AI. As for governance aspects, it is verified that public policies should be designed to create security impact reports and continuous reports on the use of AI in government systems. Internal governance measures that incorporate implementation phases are essential to achieve this. Therefore, clear roles and responsibilities should be stipulated for the ethical use of AI. In addition, it is an option to identify, correct, and address internal risks when developing AI.

In addition, EBIA proposes a Human-Centric AI approach. In simple terms, this approach places AI as a tool where human agents decide when and how to use it. At this point, implementing public policies can follow the traditional management theories concerning risk management, that is, the relationship between impact x probability. Concerning public policies in the state sector, data protection impact reports can acquire new features and increase concrete results to promote AI in conjunction with its liability and ethics. This social and multisectoral dialogue is essential to leverage the practices of accountability related to AI in organizations.

This approach can be made effective through operational management in which the entire procedure is accompanied from beginning to end, outlining its origin, purpose, data collection and processing, application, decision, regulatory models, and structural models for application. The public manager needs to know their target audience to elaborate adequate and sufficiently explainable terms of services to users. In other words, all public management processes must be based on explainability to achieve an outline of how algorithms work or even how the decision-making process was done, as well as transparency, so that information of all kinds can be obtained intelligibly and directed to the continuous development, traceability, standardization, and audit of AI practices. For this reason, communication and interaction between the parties involved are essential. More general information can be elaborated if directed to the first contact. As the citizen is faced with the effective use of AI, specific documents must be prepared to contain a step-by-step for its correct understanding, also informing those responsible for execution.

The final considerations presented here are by no means conclusions. They are only good indications for creating public policies and documents, especially for end-user feedback. The technologies that involve AI are constantly evolving and linked to ethical character, and its governance should not and cannot have fixed parameters for its application. If this happens, there would be a contradiction in the governance structure, which requires a periodic review to adapt the systems to what is intended by the user, the person in charge, and the public.

Therefore, they are initial paths to be traced and complemented by a strategy that will still have a long and **natural** way to go.

References

BARANIUK, Chris. The 'creepy Facebook AI' story that captivated the media. **BBC. 2017.** Disponível em: https://www.bbc.com/news/technology-40790258. Acesso em: 20 dez. 2021.

BECKER, Daniel; FERRARI, Isabela. O direito à explicação sobre decisões automatizadas: uma análise comparativa entre a União Europeia e o Brasil. **Revista de Direito e as Novas Tecnologias**, São Paulo, vol. 1, n. 1, out./dez. 2018.

BILEFSKY, Dan. Hackers Use New Tactic at Austrian Hotel: Locking the Doors. **The New York Times**. 2017. Disponível em: https://www.nytimes.com/2017/01/30/world/europe/hotel-austria-bitcoin-ransom.html. Acesso em: 20 dez. 2021.

BURRELL, Jenna. How the machine 'thinks': Understanding opacity in machine learning algorithms. **Big Data & Society**, [s.l.], v. 3, n. 1, p.1-12, jan. 2016.

BRASIL. Estratégia Brasileira de Inteligência Artificial – EBIA. **MCTI**. 2021. Disponível em: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/inteligencia-artificial>. Acesso em: 30 jan. 2022.

BRASIL. **Estratégia Brasileira para Transformação Digital (e-Digital).** 2018a. Disponível em: https://www.gov.br/mcti/pt-br/centrais-de-conteudo/comunicados-mcti/estrategia-digital-brasileira/estrategiadigital.pdf. Acesso em: 31 jan. 2022.

BRASIL. Manual de Gestão de Riscos. **Tribunal de Contas da União.** 2018b. Disponível em: https://portal.tcu.gov.br/planejamento-governanca-e-gestao/gestao-de-riscos/manual-de-gestao-de-riscos/. Acesso em: 30 jan. 2022.

BRASIL. **Startup Brasil**. 2022. Disponível em: https://www.startupbrasil.org.br/startups/conheca-as-startups-do-programa/. Acesso em: 30 jan. 2022.

CANADÁ. **Pan-Canadian AI Strategy**. 2017. Disponível em: https://cifar.ca/ai/. Acesso em: 30 jan. 2022.

CHINA. **A Next Generation Artificial Intelligence Development Plan**. 2017. Disponível em: https://na-production.s3.amazonaws.com/documents/translation-fulltext-8.1.17.pdf. Acesso em: 30 jan. 2022.

CITRON, D. K.; PASQUALE, F. A. The Scored Society: Due Process for Automated Predictions. **Washington Law Review**, Washington, v. 89, n. 1, p. 2-27, 2014.

CONECTA STARTUP BRASIL. **Startups**. 2021. Disponível em: https://conectastartupbrasil.org.br/startups. Acesso em: 30 jan. 2022.

COMISSÃO EUROPEIA. **Artificial Intelligence Act**. 2021. Disponível em: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206. Acesso em: 30 jan. 2022.

DIAKOPOULOS, Nicholas. Algorithmic Accountability Reporting: On the Investigation of Black Boxes. **Tow Center for Digital Journalism**, 2013.

DINAMARCA. **Strategy for Denmark's Digital Growth**. 2021. Disponível em: https://eng.em.dk/media/10566/digital-growth-strategy-report_uk_web-2.pdf. Acesso em: 30 jan. 2022

DIVINO, S. B. S. Desafios e benefícios da inteligência artificial para o Direito do Consumidor. **Revista Brasileira de Políticas Públicas**, Brasília, v. 11, n. 1, p. 655-689, 2021.

DIVINO, S. B. S.; MAGALHAES, R. A. Inteligência Artificial e Direito Empresarial: Mecanismos de Governança Digital para Implementação e Confiabilidade. **Economic Analysis of Law Review,** Brasília, v. 11, p. 72-89, 2020.

DOSHI-VELEZ, Finale; KORTZ, Mason. Accountability of AI under the law: the role of explanation. **Berkman Klein Center Working Group on Explanation and the Law**, 2017.

DUTTON, Tim. An Overview of National AI Strategies. **Medium**. 2018. Disponível em: https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd. Acesso em: 30 jan. 2022.

EL NAQA, Issam; MURPHY, Martin J. What is machine learning?. In: **machine learning in radiation oncology.** Springer: Cham, 2015. p. 3-11.

FINLAND. **Finland's Age of Artificial Intelligence**. 2017. Disponível em: https://knowledge4policy.ec.europa.eu/ai-watch/finland-ai-strategy-report_en. Acesso em: 30 jan. 2022.

FJELD, Jessica et al. Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. **Berkman Klein Center Research Publication**, Cambridge, n. 2020-1, 2020.

FLORIDI L. Soft ethics, the governance of the digital and the General Data Protection Regulation. **Philosophical Transactions of the Royal Society**, London, v. 376, n. 2133, 2018. Disponível em: http://dx.doi.org/10.1098/rsta.2018.0081. Acesso em:

FRANCE. **AI for Humanity**: French Strategy for Artificial Intelligence. 2018. Disponível em: https://super-ai.diascreative.net/ai-for-humanity-french-strategy-for-artificial-intelligence. Acesso em: 30 jan. 2022.

GAROUPA, Nuno; PORTO, Antonio José Maristrello; FRANCO, Paulo Fernando De Mello. As indenizações pela perda do tempo útil do consumidor. Espera e custos de oportunidade**. Revista de Direito do Consumidor**, Brasília, v. 124, p. 263–293, jul./ago. 2019

GARRET, Filipe. O que é algoritmo? Entenda como funciona em apps e sites da Internet. **TechTudo**. 2020. Disponível em: https://www.techtudo.com.br/listas/2020/05/o-que-e-algoritmo-entenda-como-funciona-em-apps-e-sites-da-internet.ghtml. Acesso em: 31 jan. 2022.

GLAUBITZ, Alina; RAYMOND, Nathaniel. **How should liability be attributed for harms caused by biases in Artificial Intelligence?.** 2021. Disponível em: https://politicalscience.yale.edu/sites/default/files/glaubitz_alina.pdf. Acesso em: 20 dez. 2021.

INDIA. **National Strategy For Artificial Intelligence**. 2018. Disponível em: https://indiaai.gov.in/research-reports/national-strategy-for-artificial-intelligence. Acesso em: 30 jan. 2022.

ITALY. **Strategia Nazionale per l'Intelligenza Artificiale**. 2020. Disponível em: https://knowledge4policy.ec.europa.eu/ai-watch/italy-ai-strategy-report_en. Acesso em: 30 jan. 2022.

JAPAN. **Artificial Intelligence Technology Strategy**. 2017. Disponível em: https://ai-japan.s3-ap-northeast-1.amazonaws.com/7116/0377/5269/Artificial_Intelligence_Technology_StrategyMarch2017.pdf. Acesso em: 30 jan. 2022.

KEATING, Gregory C. The theory of enterprise liability and common law strict liability. **Vanderbilt Law Review,** Nashville, v. 54, p. 1285, 2001.

KOREA. **National Strategy for Artificial Intelligence**. 2019. Disponível em: https://www.msit.go.kr/bbs/view.do?sCode=eng&mId=10&mPid=9&bbsSeqNo=46&nttSeqNo=9. Acesso em: 30 jan. 2022.

LEMLEY, Mark A.; CASEY, Bryan. Remedies for robots. **The University of Chicago Law**

Review, Chicago, v. 86, n. 5, p. 1311-1396, 2019.

MEXICO. **Artificial Intelligence Agenda MX**. 2018. Disponível em: https://oecd.ai/en/dashboards/countries/Mexico. Acesso em: 30 jan. 2022.

O GLOBO. Amazon coleta detalhes íntimos sobre seus usuários através da Alexa; veja quais. **IG Tecnologia**. 2021. Disponível em: https://tecnologia.ig.com.br/2021-11-22/amazon-alexa-coleta-dados-intimos.html. Acesso em: 20 dez. 2021.

OCDE. **Recommendation of the Council on Artificial intelligence**. 2019. Disponível em: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449. Acesso em: 30 jan. 2022.

OSOBA, Osonde; WELSER IV, William. An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence. **RAND Corporation,** Santa Mônica, 2017.

PAÍSES NÓRDICOS. **Declaration on AI in the Nordic-Baltic Region.** 2018. Disponível em: https://www.regeringen.se/49a602/globalassets/regeringen/dokument/naringsdepartementet/20180514_nmr_deklaration-slutlig-webb.pdf. Acesso em: 30 jan. 2022

PASQUALE, Frank. **The Black Box Society:** The Secret Algorithms that Control Money and Information. Cambridge: Harvard University Press, 2015. 320 p.

PEREZ, Sarah. Microsoft Silences Its New A.I. Bot Tay, After Twitter Users Teach It Racism, **TECHCRUNCH.** 2016. Disponível em: https://techcrunch.com/2016/03/24/microsoft-silences-its-new-a-i-bot-tay-after-twitter-users-teach-it-racism/. Acesso em: 20 dez. 2021.

POSNER, Richard. **Economic Analysis of Law**. 3. ed. Alphen aan den Rijn: Wolters Kluwer, 2012. 1026 p.

RACHUM-TWAIG, Omri. Whose Robot Is It Anyway? Liability for Artificial-Intelligence-Based Robots. **University of Illinois Law Reveview,** Illinois, v. 2020, p. 1141, 2020.

RUSSELL, Stuart. J.; NORVIG, Peter. **Artificial intelligence**: a modern approach. 3. ed. New Jersey: Pearson Education, 2010. 1115 p.

SANDVIG, Christian et al. Auditing algorithms: Research methods for detecting discrimination on internet platforms. **Data and discrimination: converting critical concerns into productive inquiry,** Seattle, v. 22, p. 4349-4357, 2014.

SERASA. **O que é Serasa Score 2.0?** Disponível em: https://www.serasa.com.br/score/score-2-0/. Acesso em: 03 set. 2021.

SHAVELL, Steven. **Economic analysis of accident law**. Cambridge: Harvard University Press, 2009. 352 p.

SINGAPORE. **Artificial Intelligence Governance Framework**. 2020. Disponível em: https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf. Acesso em: 30 jan. 2022.

SOUZA, Celina. **Políticas públicas: conceitos**, tipologias e subáreas. Trabalho elaborado para a Fundação Luís Eduardo Magalhães. São Paulo, 2002.

SWEDEN. **Nationell inriktning för artificiell intelligens**. 2018. Disponível em: https://www.regeringen.se/informationsmaterial/2018/05/nationell-inriktning-for-artificiell-intelligens/. Acesso em: 30 jan. 2022.

TAIWAN. **AI Taiwan Action Plan**. 2019. Disponível em: https://english.ey.gov.tw/News3/9E5540D592A5FECD/1dec0902-e02a-49c6-870d-e77208481667#:~:text=Capitalizing%20on%20this%20wave%2C%20the,greater%20momentum%20into%20Taiwan's%20industries. Acesso em: 30 jan. 2022.

UNESCO. **UNESCO member states adopt the first ever global agreement on the Ethics of Artificial Intelligence.** 2021. Disponível em: https://en.unesco.org/news/unesco-member-

states-adopt-first-ever-global-agreement-ethics-artificial-intelligence. Acesso em: 30 jan. 2022.

UNIÃO EUROPEIA. **Artificial Intelligence**: Europe needs to take a human-in-command approach, says EESC. 2017. Disponível em: https://www.eesc.europa.eu/en/news-media/press-releases/artificial-intelligence-europe-needs-take-human-command-approach-says-eesc. Acesso em: Acesso em: 11 abr. 2020.

UNITED STATES. **Artificial Intelligence Index Report 2021**. Stanford University. 2021. Disponível em: https://aiindex.stanford.edu/wp-content/uploads/2021/11/2021-AI-Index-Report_Master.pdf. Acesso em: 30 jan. 2022.

UNITED KINGDOM. **AI Sector Deal.** 2019. Disponível em: https://www.gov.uk/government/publications/artificial-intelligence-sector-deal/ai-sector-deal. Acesso em: 30 jan. 2022.

VINCENT, J. Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day. **The Verge.** 2016. Disponível em: https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist. Acesso em: 20 dez. 2021.

WANG, Ge. **Humans in the Loop***:* The Design of Interactive AI Systems. Human-centered Artificial Intelligence. Stanford University. 2019.

WIPO. **Índice Global de Inovação.** 2021. Disponível em: https://www.wipo.int/edocs/pubdocs/pt/wipo_pub_gii_2021_exec.pdf. Acesso em: 30 jan. 2022