

Artificial Intelligence Regulation: A Meta-Framework for Formulation and Governance

Patricia G. Rêgo de Almeida
Universidade de Brasília
Câmara dos Deputados
patricia.almeida@camara.leg.br

Carlos Denner dos Santos
Universidade de Brasília
Université du Québec-Montreal
carlosdenner@unb.br

Josivania Silva Farias
Universidade de Brasília
josivania@unb.br

Abstract

This article presents a meta-framework for Artificial Intelligence (AI) regulation that encompasses all stages of international public policy-making, from formulation to sustainable governance. Based on a vast systematic review of the literature on Artificial Intelligence Regulation (AIR) published between 2009 and 2019, a dispersed body of knowledge organized under the label “framework” was identified, containing 15 unique frameworks and several different theories that created a complex scientific scenario for research and practice. Theories and principles as diverse as Agile and Ethics were found. Thus, a structured analytical method was followed to integrate this bulk of knowledge into a cohesive, synthetic, and generic theoretical tool. The resulting “AIR framework” provides a trustworthy lens for societies to think collectively and make informed policy decisions related to what, when, and how the uses and applications of AI should be regulated. Moreover, the novel framework organizes the latest developments in the area in a format that allows future research to be framed in and added to the published literature. The (potential) impacts of AI on society are immense, and therefore the discourses, social negotiations, and applications of this technology should be guided by common grounds in terms of terminology, governance, and social values.

1. Introduction

The widely disseminated use of AI in our daily actions and in an unnoticeable fashion [1] has introduced unprecedented legal issues in exceptional concepts and scenarios [2].

From the same perspective, the opacity of data processing in a machine learning solution increases the likelihood of legal surprises [3].

Based on this reflection, this work sought to carry out a vast search for literature that is relevant in terms of Artificial Intelligence Regulation, processing and grouping it into an integrative theoretical framework that allows for reflections and actions aimed at regulating operations and relationships between both natural and legal persons and systems with embedded AI.

2. Reasons to Regulate AI

The responsibilities, security, intellectual property, and privacy associated with different systems for medical robots, drones, autonomous cars, among several “intelligent solutions” offered every day have been questioned. Illustrating the level of risk-related indeterminacy, machine learning has been combined with game theory [4] in cases where developers were using game theory to help teach strategic defense to algorithms. A game between two algorithms predicted that one would kill the other only when there was an absolute scarcity of resources. However, when a more intelligent algorithm was introduced, it immediately killed the weaker ones [5]. This case reinforces the idea that an autonomous system will inevitably find itself in a situation in which it needs not only to obey a certain rule or not, but also make a complex ethical decision [6].

Considering all those risks, establishing best practices for delegating and defining new moral responsibility attribution models is crucial in order to leverage the opportunities created by AI [7]. Risk assessment models can provide support and flexibility to big data and AI applications [8]. Bestowing a sense of morality upon this superintelligence should be a priority, despite the difficulties associated with putting that into practice [9]. From an ethical and moral perspective, a decision is deemed acceptable insofar as

it does not violate the principles of the ethical framework [10].

The reasons to regulate include: manufacturers' need to comprehend a legal framework within which they can operate reliably; consumers' and society's need to be protected from devices that may harm or adversely affect them; and the need for business opportunities [11].

In industries still lacking regulation, the general approach observed is that innovation is freely allowed, but those in charge should bear the consequences in case certain types of damages are caused [12].

3. Seeking the Best Way to Regulate

When used to denote an attempt to standardize behavioral patterns, the term "regulation" assumes the meaning of a law [13]. At the present, the few existing laws are resorted to in order to judicially settle damages brought about by AI-supported products and services. If, on the one hand, cases are multiplying, on the other, the legislative branch seems to be moving at a negligible speed compared to the technological advancements [14][1].

A still unsolved equation is the breadth of laws dealing with globally produced and commercialized technologies [11] and robot-generated inventions [15]. The problem reaches even broader dimensions when one considers the complex networks established in the technology industry, making it possible for products to be subjected to learning from data distributed all over the world [16].

Large-scale data analyses have revealed that the key challenge related to the AI regulation dilemma is demonstrating it is produced and deployed appropriately [3]. One of the most advocated strategies is transparency, an opening of the entire production process, especially the decision-making rules, the method, and the basis utilized when training the intelligent system [17][3][18]. In certain situations, the regulation will have to be enforced through algorithms. Thus, an autonomous system would have guardian algorithms to ensure the parameters are within predefined standards [5]. A similar strategy to open data is the Explainable Artificial Intelligence (XAI) standard for the creation of coding models oriented towards a global comprehension [7][19].

Some of the theories that have been proposed to regulate AI are based on contractual and extracontractual liability or on strict liability and adopt an irreproachable liability model in the case of AI, since the moral responsibility is distributed among designers, regulators, and users. The attempt to hold robots accountable for their actions has led a few countries to consider the possibility of granting a legal

identity to each unit. One could argue that if parties in a contractual relationship may be legally represented by another entity, then so can systems [20]. As a counterargument, the term "robot liability" should be replaced with "indirect liability over the robot", given the impossibility of claiming damages from a robot; therefore, it cannot be held criminally liable. Hence, the impact of such products on society should also be a liability [21][22].

Also among the concerns that motivate AI regulation is the approach aimed at minimizing the disruption of the work model with the goal of fighting job loss [23].

Drawing attention to the domain of what is to be regulated, attempts to legislate on digital technologies without proper knowledge for doing so have been criticized [12]. With the intention of minimizing those risks, a gradual regulation strategy [14] can be used. When mitigating risks, regulatory agencies could bar the introduction of certain algorithms into the market until their safety and efficacy have been proven by means of tests [17] founded on ethics [24].

A milestone was reached with the General Data Protection Regulation (GDPR), which fights discrimination and reinforces the right to explain a decision made based on smart algorithms [25]. In 2017, the European Parliament Committee on Legal Affairs released a report recommending the creation of a European agency for robotics and AI, and suggests a combination of hard and soft laws, given the complexity associated with the evolution of the regulatory model [26]. Another highlight in European legislation were the reports released by the House of Lords [27] which underscores the need to create a regulatory framework [26]. Another effort observed in the U.S. resulted in H.R. 4625 [28], which seeks to define the conditions for utilizing and commercializing AI through the establishment of the Federal Advisory Committee on the Development and Implementation of Artificial Intelligence. Additionally, several countries have shown their intention to create policies and laws to regulate the development and use of AI [29].

4. Method

With the goal of surveying the international debate on AI regulation found in the literature, we systematically searched for and cataloged articles to compile the bibliometrics and perform a qualitative analysis to demonstrate the evolution of said debate as a basis for any future regulation efforts.

We opted to gather materials published between 2009 and June 2019, searching by title and subject in the ScienceDirect, JSTOR, SpringerLink, PROQUEST, IEEE, Scopus, DOAJ, and Google

Scholar databases. Only peer-reviewed research articles in English were collected, while dismissing duplicates. The selection was refined by reading all subjects with the goal of removing *locus* discussion cases from the sample, as well as those in which regulation was not the main topic under discussion.

The sample was sorted according to specific parameters when structuring the demographics: year of publication, journal, author, author's institution, author's field of study, country, keywords. We also wrote down for each article: concepts, findings, contributions, agenda, approach, method and researched subject. The following terms were considered when classifying the subject: "risks", "ethics", "how to regulate", "existing regulation", "framework". After an analysis of the abstracts, a sample comprising 51 articles was selected for further reading and discussion.

5. Results

In terms of timeline, it is worth highlighting that 94% of the articles were published after 2015, with a growing production every year thereafter. The exclusively qualitative approach was dominant, being observed in 47 articles, whereas the mixed approach was found in only 4. This is to be expected when dealing with such an incipient topic. The initial debate is exploratory in this case, which explains the substantial number of works that are still qualitative.

The sample reflects the evolution in the fields of research that take an interest in AI regulation. Although Artificial Intelligence as a subject of study traditionally pertains to Computer Science (IT) and Engineering, there has been a growing interest in its regulation by other areas, such as Law, Business Administration, and Philosophy. Out of the entire sample, researchers from the field of Law represent 53%, followed by IT (43%). In some cases, the same article is coauthored by researchers from different areas.

Special attention was paid to the analysis of the main object of the sample, non-exclusively divided into: "Risks" (41%), "Ethics" (16%), "How to Regulate" (65%), "Existing Regulation" (8%) and "Framework" (26%). It is worth noting that concerns over risks and ethics as applied to AI have been a constant with the passing of the years. Yet, discussions on how to regulate only became significant in 2016. With regard to the discussions on AI regulatory frameworks, the largest concentration occurred after 2017, adding up to 15 proposals found in the samples, which will be presented and analyzed next.

5.1. Model for Ethical Issues in Experimental Technologies [30]

Based on the premise that a robot is an experimental technology, this model intends to minimize the ethical dilemmas associated with decisions made by autonomous systems [31]. The proposal supports decision-making processes based on 16 conditions for deploying experimental technologies built to anticipate potential ethical issues as robots interact with people and the environment. Split into three groups, the conditions are aimed at: preventing damages (non-maleficence conditions), good-doing (beneficence conditions), and respect for autonomy and justice. Concerns over the risks extend to the prediction of "red button" conditions. They also recommend implementing this model as part of a gradual interactive strategy.

5.2. Interactive Regulatory Governance Model [14]

The proposal is based on an interactive governance model for technological development and law formulation processes in which the attributions of stakeholders are highlighted. The need for continuous learning and a gradual evolution of the legal framework is noteworthy, using the expressions "Regulatory Innovation" and "Temporary Experimental Legislation", and considering the proper sequence of actions among agents at the maturity stage of an innovation's lifecycle. The proposed model includes components such as:

- A Regulatory-to-Technology (R2T) macroprocess to guide the creation of a new conceptual model for robots pursuant to the existing legislation.
- A Technology-to-Regulatory (T2R) macroprocess to adjust the law to the needs resulting from technological evolution.
- A data repository shared by R2T and T2R.

Among the main benefits of this hybrid AI Governance Model, it is worth highlighting the integration of top-down with bottom-up regulatory actions in an incremental strategy, thus minimizing the risk posed by regulating a new, constantly changing object.

5.3. Ethics Model for AI Development and Deployment [32]

Founded on philosophical principles and the dimensions of maintaining human rights and well-being, the proposed ethical framework for AI development and deployment is divided into: Ethical Perspectives - Rights (deontological ethics); Damages

and Goods (teleological ethics); Virtue (aretaic ethics); Community (community ethics); Dialog (communication ethics); and Flourishing (flourishing ethics) – for the core functions of AI. These core functions are considered as being the following: identification of ethical issues, development of human consciousness, collaborative engagement, liabilities and integrity of AI.

5.4. Competency-Based AI Regulatory Model [33]

Considering the competencies, strengths and weaknesses of each state power, the proposal of an AI Regulatory Model based on the distribution of responsibilities without losing sight of the mission goals. The model acknowledges the regulatory agencies of the executive, legislative and judicial powers as agents in the regulatory process.

In the proposed model, the legislative branch would provide a statute putting the regulatory agencies in charge of certifying products and services that use AI in terms of user and social safety. Supported by groups of researchers, regulatory agencies would be more agile and competent to monitor the technological evolution, to identify risks in the intelligent learning process and AI utilization, to issue technical recommendations, as well as to verify whether the technology is being applied for its declared purposes. If a company's products or services cause any damages, if certified, the company would be judged based on more lenient rules, whereas uncertified companies would be subjected to more rigid rules. Courts would judge companies for any losses and damages caused, considering the situation in which those organizations find themselves in the context of certification.

5.5. Regulatory Model Sustained by Society [34]

Inspired by the Social Contract Theory [35], the Regulatory Model Sustained by Society adjusts the “man-in-loop” to the “society-in-loop” model.

The agility and effectiveness of the interactive learning machine (man-in-loop) stems from users feedbacks, thus enriching the generated knowledge. If used to learn problems resulting from the use of AI in products and services, society-in-loop would become a governance tool for society to control and proactively identify those elements. Conflicts among safety-, privacy-, and justice-related concepts would benefit from this model. This relationship can be summed up as: society-in-loop = man-in-loop + social contract.

5.6. Principles of Robotics [36]

Highlighting the responsibilities of all agents involved in robotics, five principles were established for robot designers, manufacturers and users. The main goal of the rules is to emphasize that robots are tools, whereas humans are the actual responsible agents.

The opportunity to use this proposal in audits performed by regulatory agencies can be identified, and that need must be reflected on the legislation to be adapted or created.

5.7. Agile AI Governance [37]

This would be an alternative to the problem of temporal mismatch between formal regulatory actions and the production and commercialization of deep machine learning-based products and services. The success of this proposal depends on the amount of effort put into it by the market, scholars, government, insurance companies, and organized civil society. The model predicts actions performed by a Governance Coordinating Committee and a Global Governance Coordinating Committee. The international approach is also advocated as a means to provide some balance to the several countries that are not yet participating in the AI regulation dynamics, considering that the current situation makes them more vulnerable. This model is a soft law that mitigates risks while the legislation is drawn up. The soft governance part involves industry standards, social codes, labs, certification practices, procedures, and programs. The hard governance part concentrates laws, regulations, and regulatory groups.

The proposed model takes a relationship network into account in order to address AI in a way that bolsters the formulation of actual standards while the legislation matures.

5.8. Sustainable AI Development [38]

Concerns over the entire lifecycle of an AI-based solution were the main foundation considered when devising the Sustainable AI Development (SAID) framework. Analyzed under the lens of a governance structure, SAID is stratified into the following layers: Technological (data, architecture and algorithm design), Social (analysis of the potential consequences of using AI in the social sphere) and Governance (the way algorithms influence both national and international decisions).

5.9. Ethical Framework for Automations that Use Robotics [23]

Concerned with the integration of several stakeholders with automations using AI, this framework integrates the Stakeholders Theory with the Social Contract Theory in an attempt to find ethical grounds for the use of AI. The proposal considers stakeholders as being: workers, the market, governments, the economy, and society in general. The impacts on the job market and new actions and relationships among those stakeholders are greatly emphasized. The framework follows a set of steps that goes from the identification of stakeholders to an analysis of social contracts, impact assessment, and lastly, actions aimed at mitigating the risk of terminating or breaching work contracts.

It is worth noting that this is the only proposal that considers as stakeholders those workers whose jobs or occupations will be modified with the introduction of AI into products and services.

5.10. Intelligent Model to Regulate Learning Algorithms [18]

Focused on a strategy to fight intelligent services that contain biases, this model proposes that an algorithm should assess the basic elements of a machine learning process (data, testing algorithms, and decision models). The proposal is founded on the thesis that the transparency of a code is insufficient to guarantee an unbiased solution and admits that even when learning from vast amounts of data, it is still possible to find biases. It also recognizes the difficulty to identify those problems automatically as algorithms grow in complexity. The study then goes on to analyze the characteristics of an algorithm that could classify it as capable of detecting bias-related issues in a learning process.

5.11. Universal Declaration of Human Rights as a Framework [39]

This model is founded on the argument that the several different frameworks related to each specific area of ethics are insufficient to regulate AI on an international scale, both in private sectors and within the government. Due to that gap, the Universal Declaration of Human Rights [40] was considered as a necessary approach to the effective regulation of AI according its impact on society.

5.12. Software Requirement Model for the Ethical Assessment of Robots [41]

The proposal puts forth a set of general specifications to be considered in a system aimed at

assessing robots during their construction. Different elements are taken into account in the suggested specifications, such as the user's emotional state.

It seems the proposal may be utilized by the industry and regulatory agencies alike. In both cases, it could be the first red flag signaling the need for a red button in robot projects [42].

5.13. Ethical Judgement Model for Codes [43]

Considering that tackling ethical decisions is better than avoiding them, the author proposes a formal logical model that can be implemented in an agent facing an ethical dilemma with the ability to both make decisions and explain those decisions.

The concepts of 'decision', 'event', and 'effect' were taken into account when building the model's functionalities. Ethical framework principles were also gathered – Consequentialist Ethics, Deontological Ethics, and the Doctrine of Double Effect, formalized in judgment functions that return three possible results: acceptable (\top), unacceptable (\perp), or undetermined (?).

5.14. Asilomar AI Principles [44]

The governance model proposed by the Asilomar Conference resulted in 23 AI Principles undersigned by thousands of experts [45]. Grouped under 'Research Issues', 'Ethics and Values', and 'Longer-Term Issues', those principles encompass the lifecycle of an AI-embedded product or service – from motivation and funding to the assessment of benefits and judgement criteria concerning its impacts.

5.15. European Ethics Guidelines for Trustworthy AI [46]

With the goal of creating guidelines to orient a new AI Governance, the European Commission, through a group of experts, has drawn up the Ethics Guidelines for a Trustworthy AI, based on a very comprehensive structure and divided into three tiers [45]. The highest tier addresses four ethical principles founded upon fundamental human rights. The second tier includes seven Key Requirements that are necessary for the application and ongoing assessment of an AI-based system or service throughout its lifecycle. For the base tier, a list of recommendations directed at the operationalization of the key requirements in the upper tier for each specific system has been formulated.

6. Frameworks Approaches

An analysis of the approaches adopted by each of the 15 frameworks proposed in the sample resulted in Table 1.

The fact that Ethical Guidelines exist is not enough to have any effect on the software development industry. Thus, models that are strongly grounded on ethical principles require legal mechanisms so that those recommendations can be fulfilled [47].

Table 1 – Comparative table of the approaches explored in the frameworks, compiled by the author.

Approach	[30]	[14]	[32]	[33]	[34]	[36]	[37]	[38]	[23]	[18]	[39]	[41]	[43]	[44]	[46]
Institutional Competences	■		■				■								■
International							■	■			■				■
AGILE Thinking							■								
Hybrid (soft + hard)		■		■			■								■
Successive Interactions		■													■
Regulatory Agency		■		■			■								■
Chronology of the Regulation	■	■		■			■	■							■
Gradual Improvement	■	■		■	■		■								■
Ethical Principles	■		■			■			■		■	■	■	■	■
Social Contract					■				■						
Job Market									■						
Impacts on Stakeholders		■	■		■		■	■		■		■		■	■
Governance		■	■	■			■	■	■	■		■			■
Process Definition		■							■	■					■
Technology as a Regulator											■	■	■		

Frameworks that take the social contract into account rank among the most open to the participation of society in a coproduction with the government. Those models view citizens as outstanding stakeholders. Concerns over the impacts on the job market are also a way to assess the impact on stakeholders.

The advocacy of a gradual deployment of the regulation is a risk-mitigation strategy, but it could also be combined with successive interactions between the legislative branch and the regulatory agencies, thus enabling continual improvement during the legislative process.

The Interactive Regulatory Governance Model, the Competency-Based Regulatory Model, the Agile Governance, the Asilomar Principles and the European Ethics Guidelines for Trustworthy AI proposals encompass a larger number of topics. The European proposal highlights that a trustworthy AI must be lawful, ethic and robust. The others explore the relationship among all parties involved in the regulation process and the attempt to find balance between the more rigid and the more flexible mechanisms. It is worth noting that the Agile

Governance proposal does not exclude conventional actions for a formal regulation – the Interactive Regulatory Governance Model and the Competency-Based Regulatory Model, both of which involve the legislative branch. Therefore, this configures a transitional situation in which consensual standards would be agreed upon and enforced, and the risks would be mitigated while legal mechanisms are not made official, which is very similar to the concept of Dynamic Regulation, in which feedback serves as a basis for the maturity of the regulatory instrument [48].

The relationship between the proposal put forth by AI4PEOPLE [49] and the presented frameworks cannot go unnoticed. When analyzing several movements advocating the establishment of criteria for how to best use AI, studies identified an opportunity to develop a competition around a technological reform [50].

The scope of these actions encompasses stakeholders more comprehensively than the models in the sample that raised such concerns. Pondering over the need to find synergy among global AI regulation-oriented actions, in the AI Agile Governance [37], the creation of a Governance Committee for each country and a Global Governance Committee in an international context was suggested, which was also touched upon in two articles included in the sample.

Despite the small number of existing software-based regulation models, similar models are likely to arise, since the increasing complexity of AI solutions results in more system rules [51][52][53][54], which in turn means a higher probability of conflicts between those rules in combined systems [55]. A problem, therefore, that exceeds the human capacity to follow.

7. AI Regulatory Meta-Framework

The supplementary nature of some frameworks confirms the perception that the impacts of AI and robotics would demand a combination of design, laws, and education [56]. When arguing that a framework is insufficient to address such a multidisciplinary topic [10] embedded into the political and societal context [57], an AI Regulatory Meta-Framework was built to include the main contributions from each model in the examined sample (Fig. 1).

The Government’s exclusive competencies would be distributed across the legislative, the executive, and the judicial branches.

Apart from making laws, it is important to maintain the legislative branch open so that its bills can be discussed with society and academia (B), receiving constant feedback (F). Strongly represented by a regulatory agency created by the legislative through a statute (J), the executive branch would then establish a

relationship with the legislative as part of an ongoing process, in which the legislative would survey the impact on the legislation and its evolution based on the knowledge obtained from the regulatory agency (T2R), much like the regulatory agency structures its internal

work processes based on the legislation discussed and approved by the legislative (R2T). The quality and efficiency of this synchronicity between T2R and R2T processes are strengthened through a database shared by the legislative and the regulatory agency.

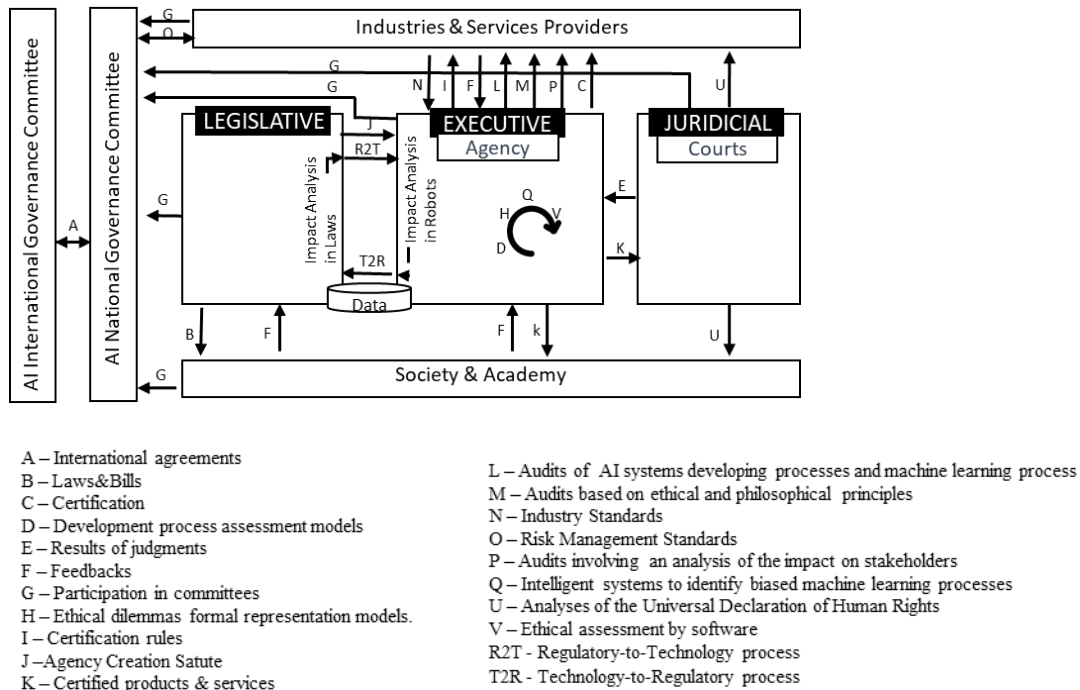


Fig 1 – AIR Framework

Among the regulatory agency’s competencies, the creation and application of models to assess the development and learning processes of AI systems (D) stand out. As in the legislative branch, an open practice by the regulatory agency is likewise desirable, receiving feedback from society and academia alike (F). Companies that submit their products to the regulatory agencies, after a successful appraisal, would receive a certificate (C) within their field of action (transport, healthcare, entertainment, education, military, etc.). The strictness and nature of the assessment processes could be different for each of those fields. The issuance of certificates could be a strategy to be applied before laws are passed, since they already inform society, in a transparent fashion, about the safety levels and risks of the products and services it consumes. Advertising campaigns by the government and certified companies would also strengthen that strategy. Law enforcement by courts would also undergo a continuous learning process with regard to interpretations based on the legislation in effect, as well as on new laws. In countries where the certification is incorporated into laws, decisions on cases involving uncertified companies would be treated

differently from those involving certified companies. Thus, courts would need to have up-to-date information on each company’s certified products and services (K). Considering a continuous learning process, the regulatory agency would receive the results of decisions involving AI systems (E), which would then be stored in the database shared with the legislative branch.

Through a quick process, industries and service providers would need to receive the regulatory agency’s certification rules stated as clearly as possible (I), while providing feedback (F) on the conditions that preclude the development process required by the regulatory agency from moving forward.

The audit conducted by the regulatory agency would take place in three dimensions. Firstly, an audit of the ethical principles (M), which would also include the Universal Declaration of Human Rights and the related legislation in each country. A second dimension would occur through an audit aimed at assessing the impact on stakeholders (P), even when no ethical issues or dilemmas are entailed. Through this analysis, future problems arising from new arrangements made by society could be identified. Failures in the basic

elements of trusted AI system would also be identified (transparency, privacy, human wellbeing, accountability, etc.). And lastly, an audit of the technical procedures followed when building the AI system or of the learning process to which it was subjected (L).

The efficiency and knowledge expected by the regulatory agency depend on: formal representation models for ethical dilemmas with functions designed to solve them (H); systems to identify biased learning processes (Q); systems to evaluate ethics in robot actions (V) and Development process assessment models (D). In courts, the Universal Declaration of Human Rights would become the foundation for interpreting various situations that are not yet regulated by law or do not need to be treated on a legal level.

On a national level, discussions to facilitate priority actions and the recognition of industry standards would be enabled through an AI Governance Committee, bringing together government agencies and industry representatives, service providers, and scholars (G). The agreed upon standards (N) make it possible to move forward in some technological dimensions while the legislative discusses adjustments to the legislation. The risk management criteria (O) related to the use of those standards would be negotiated between the national committee and the industry.

The plethora of components in AI services and products of global reach imposes actions that would be agreed upon in an International Governance Committee comprising representatives from each country's committee (A). On many occasions, transparency in production processes is only feasible through international agreements.

8. Conclusion

The need and urgency to regulate Artificial Intelligence both in Brazil and worldwide seems indisputable. The complexity of the topic is also evident, whether due to the advanced nature of technology or because its impacts structurally affect social standards.

A study of the literature by means of a sample comprising 51 articles published between 2009 and 2019 revealed significant efforts to identify and scale the risks and ethical dilemmas related to AI, and also to seek a model for regulating AI through different modalities, which is being monitored by governments.

We realized the birth of a reshaping of the perception of the law, as how occurred with disruptive innovations in the past [58]. The heterogenous nature of the professional profiles involved in the debate evinces the complexity and maturity with which the

topic is being studied. Such an in-depth approach, on the one hand, may have caused certain delays in research, but on the other, it has prevented inappropriate regulatory solutions from being made official.

The discussed frameworks are based on supplementary approaches and therefore are insufficient when analyzed separately. The consolidation approach proposed as a Meta-Framework (Fig. 1) seems to be the most adequate strategy for the deployment of an AI governance, given the existence of several agents and the laterality of the topic, intertwining different areas of knowledge. The expanded view of the presented AIR Framework will enable the involved agents to identify their role, while establishing a roadmap for a gradual, uninterrupted deployment.

In addition to this, it will contribute for the creation of a new model of rewards and punishment to balance this new reality [59][60] and taken into account the world as it will be [61].

On the path to improve each component of the AIR Framework, more than bringing them closer together, there needs to be a synchronization of the agents involved towards a sustainable regulation. Along that journey, an alliance between scholars and the government's three agents (the executive, legislative and judicial branches) is crucial to the regulation macroprocess.

The countries leading the debate are probably ready to arrange the partnerships and agreements among institutions that are necessary for a comprehensive and effective governance, as well as to initiate a regulation process. Nonetheless, the release of products with embedded AI in countries that have advanced regulation models, in and of itself, does not guarantee the same safety levels for countries that are still unripe in this regard.

Much is yet to happen in the formulation of solutions using frameworks in real-case scenarios so as to enable an empirical analysis and studies for the evolution of the frameworks presented in the examined sample, as well as for the improvement of the proposed Meta-Framework, thus culminating in the creation of a reference model of AI governance in which maturity levels would be established, which could be monitored by international bodies in a collaborative effort. The way we and future generations live our lives depend on that cooperation.

9. References

- [1] Cerka P, Grigiene, J., & Sirbikite, G. "Liability for damages caused by artificial intelligence" *Computer Law & Security Review*, 31(3), 2015, pp. 376-389.

- [2] Calo, M. R. "Robotics and the Lessons of Cyberlaw" *California Law Review*, 103, (3), 2015, pp. 513-63.
- [3] Butterworth, M. "The ICO and artificial intelligence: The role of fairness in the GDPR framework" *Computer Law & Security Review*, 34, , 2018, pp. 257-268.
- [4] Conitzer, V., Sinnott-Armstrong, W., Borg, JS, Deng, Y. and Kramer, M. "Moral Decision Making for Artificial Intelligence" *AAAI Publication, 31^o Conference on Artificial Intelligence*, 2017.
- [5] Firth-Butterfield, K. "Artificial Intelligence and the Law: more questions than answers" *Scitech Lawyer*, 14, 2017, pp. 28-31.
- [6] Dennis, L., Fisher, M., Slavkovik, M. and Webster, M. "Formal verification of ethical choices in autonomous systems" *Robotics and Autonomous Systems*, 77, 2016, pp. 1-14.
- [7] Taddeo, R. and Floridi, L. "How AI can be a force for good: An ethical framework will help to harness the potential of AI while keeping humans in control" *Science Review*, 361(6404), 2018, pp. 751-752.
- [8] Mantelero, A. "AI and Big Data: A blueprint for human rights, social and ethical impact assessment" *Computer Law & Security Review*, 34(4), 2018, pp. 754-772.
- [9] Davis, E. "Ethical guidelines for a superintelligence" *Artificial Intelligence Review*, 220, 2015, pp121-124.
- [10] Bonnemais, V., Saurel, C. & Tessier, C. "Embedded ethics: some technical and ethical challenges". *Ethics and Information Technology*, 20, 2018, pp. 41-58.
- [11] Holder, C., Khurana, V., Harrison, F. & Jacobs, L. "Robotics and law: Key legal and regulatory implications of the robotics age (Part I of II)" *Computer Law & Security Review*, 32(3), 2016, 383-402.
- [12] Reed, C. "How should we regulate artificial intelligence?" *Philosophy Transactions of the Royal Societ*, 376, (2128), 2018.
- [13] Hildebrandt, M. "Algorithmic regulation and the rule of law" *Philosophy Transactions of the Royal Society*, 376 (2128), 2018.
- [14] Villaronga, E.F. & Heldeweg, M. "Regulation, I presume? said the robot – Towards an iterative G regulatory process for robot governance" *Computer Law & Security Review*, 21 June, 2018.
- [15] Holder C., Khurana V., Hook J., Bacon G. and Day R. "Robotics and law: key legal and regulatory implications of the robotics age (Part II of II)" *Computer Law Secure Review* 2016; 32:557–576.
- [16] Lenardon, J.P. A.. "The Regulation of Artificial Intelligence" *Master Thesis. Tilburg Institute for Law, Technology and Society*. Netherlands, 2017.
- [17] Tutt, A. "An FDA for Algorithms". *Administrative Law Review*, 69(83), 2017, pp. 83-123.
- [18] Buiten, M. "Towards Intelligent Regulation of Artificial Intelligence" *European Journal of Risk Regulation*, 10(1), 2019, 41-59.
- [19] Adadi, A. and Berrada, M. "Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI)" *IEEE Access Review*, September, 6, 2018, pp. 52138-52160.
- [20] Cerka, P., Grigiene, J & Sirbikyte, G. "Is it possible to grant legal personality to artificial intelligence software systems?" *Computer Law & Security Review*, 33(5), 2017, pp. 685-699.
- [21] Neverjans, N. "European Civil Law Rules in Robotics. European Parliament", *Policy Department Citizens' Right and Constitutional Affairs*, 2016
- [22] Jackson, R.W. "Artificial Intelligence and the Fog of Innovation: a deep-dive on governance and the liability of autonomous systems", 35 *Santa Clara High Tech. L.J.* 35, 2019.
- [23] Wright, S.A. & Schultz, A. "The rising tide of artificial intelligence and business automation: Developing an ethical framework" *Business Horizons*, 61(6), 2018, pp. 823-832.
- [24] Arkin, R.C. "Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture" *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, 2011, pp 121-128.
- [25] Goodman, B. and Flaxman S. "European Union regulations on algorithmic decision-making and a "right to explanation". *Artificial Intelligence Review*, 38, (3), , 2017, pp. 50-57.
- [26] Cath, C., Watcher, S., Mittelsadt, B., Taddeo, M. and Floridi, L. "Artificial Intelligence and the 'Good Society': the US, EU, and UK approach", 2017. Available at: <https://ssrn.com/abstract=2906249> or <http://dx.doi.org/10.2139/ssrn.2906249>
- [27] House of Lords, "AI in the UK: ready, willing and able?", *Select Committee on Artificial Intelligence*, Report of Session 2017-19. 13 March 2018.
- [28] House of Representatives. H.R. 4625 – 115 TH Congress (2017-2018) – Future of Artificial Intelligence Act of 2017.USA.
- [29] Future of Life Institute "National and International AI Strategies" Available at <https://futureoflife.org/national-international-ai-strategies/>
- [30] Amigoni F. and Schiaffonati, V. "Ethics for Robots as Experimental Technologies" *IEEE Robotics & Automation Magazine*, March 25, 2018, pp. 30-36.
- [31] Poel, I. V. "An ethical framework for evaluating experimental technology" *Sci. Eng. Ethics*, vol. 22, no. 3, 2016, pp. 667–686.
- [32] Schrader, D. & Ghosh, D. "Proactively Protecting Against the Singularity: Ethical Decision Making AI" *IEEE*

- Computer and Reliability Societies Review*, 16(3) , 2018, pp. 56-63.
- [33] Scherer, M.U. “Regulating Artificial Intelligence Systems: Risks, Challenges, Competences and Strategies” *Harvard Journal of Law & Technology*, 29(2) , 2016, pp. 354-398.
- [34] Rahwan, I. “Society-in-the-loop: programming the algorithmic social contract” *Ethics and Information Technology*, 20, 2017, pp. 5-14.
- [35] Rousseau, J, “The Social Contract, available at <https://www.britannica.com/topic/The-Social-Contract>.
- [36] Boden, M.; Bryson, J.; Caldwell, D.; Dautenhahn, K; Edwards, L.; Kember, S.; Newman, P.; Parry V., Pegman, G.; Rodden, T.; Sorrell, T.; Wallis, M.; Whitby, B. and Winfield, A. “Principles of robotics: regulating robots in the real world” *Connection Science*, 29:2, 2017, 124-129.
- [37] Wallach, W. & Marchant, G.E. “An Agile Ethical/Legal Model for the International and National Governance of AI and Robotics” *Association for the Advancement of Artificial Intelligence*, 2018. Available at <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6191666/>.
- [38] Djeflal, C. “Sustainable AI Development (SAID): On the Road to More Access to Justice” December 10, 2018. Available at <https://ssrn.com/abstract=3298980> or <http://dx.doi.org/10.2139/ssrn.3298980>
- [39] Donahoe, e. and Metzger, M. M. “Artificial Intelligence and Human Rights” *Journal of Democracy* 30(2), Johns Hopkins University Press, Retrieved Jun 12, 2019, 115-126.
- [40] Kunz, J. “The United Nations Declaration of Human Rights” *American Journal of International Law*, 43(2), 1949. 316-323.
- [41] Millar, J “An Ethics Evaluation Tool for Automating Ethical Decision-Making in Robots and Self-Driving Cars”, *Applied Artificial Intelligence*, 30(8), 2016. pp. 787-809.
- [42] Arnold, T. & Scheutz, M. “The big red button is too late: an alternative model for the ethical evaluation of AI systems” *Ethics and Information Technology*, 20, 2018, pp. 59-69.
- [43] Bonnemains, V., Saurel, C. & Tessier, C. “Embedded ethics: some technical and ethical challenges”. *Ethics Inf Technol.* 2018. 20: 41.
- [44] Future of Life Institute. “Ansilomar AI Principles” available on <https://futureoflife.org/ai-principles/>. Accessed in 20/09/2019.
- [45] Kosuka, S. “A governance framework for the development and use of artificial intelligence: lessons from the comparison of Japanese and European initiatives.” *Unif. L. Rev.*, Vol. 24, 2019, 315–329.
- [46] European Commission. “Ethics Guidelines for Trustworthy AI”. High Level Expert Group on Artificial Intelligence. 2019
- [47] Hagendorff, T. “The Ethics of AI Ethics - An Evaluation of Guidelines” *CoRR, abs/1903.03425*, 2019.
- [48] Kaal, W. A. & Vermeulen, E. P.M. “How to Regulate Disruptive Innovation - From Facts to Data” *Jurimetrics*, Volume 57, Issue No. 2, 2017.
- [49] AI4People “Ethical framework for a good society: opportunities, risks, principles, and recommendations.” *Atomium – European Institute for Science, Media and Democracy*. Available at <http://www.eismd.eu/wp-content/uploads/2019/02/Ethical-Framework-for-a-Good-AI-Society.pdf> accessible in 21/06/19.
- [50] Greene D et al. “Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning” *Hawaii International Conference on System Sciences* 52nd, 2019.
- [51] Verheij, B. “Formalizing value-guided argumentation for ethical systems design” *Artificial Intelligence & Law Review*, 24(4), 2016, pp. 387-407.
- [52] Prakken, H. “On the problem of making autonomous vehicles conform to traffic law” *Artificial Intelligence & Law*, 25, 2017, pp. 341-363.
- [53] Liu, H. “Irresponsibilities, inequalities and injustice for autonomous vehicles” *Ethics Information Technology Journal*, 19, 2017, pp. 193-207.
- [54] Lamo, M. and Calo, R. “Regulating Bot Speech” *UCLA Law Review* 2019, July 16, 2018.
- [55] Bench-Capon, T. & Modgil, S. “Artificial Intelligence & Law Review”, 25, 2017, pp. 29-64.
- [56] Calo, M.R. “Peeping Hals” *Artificial Intelligence Review*, 175, 2011, pp. 940-94.
- [57] Leitner C. and Stiefmueller C.M. “Disruptive Technologies and the Public Sector: The Changing Dynamics of Governance” In: Baimenov A., Liverakos P. (eds) *Public Service Excellence in the 21st Century*. Palgrave Macmillan, Singapore, 2019, pp 238-239.
- [58] Gurkaynak, G., Yilmaz, I. and Haksever, G. “Stifling artificial intelligence: Human perils” *Computer Law & Security Review*, 32(5) , 2016, pp. 749-758.
- [59] Waser, M. “Designing, Implementing and Enforcing a Coherent System of Laws, Ethics and Morals for Intelligent Machines (including Humans)”, *Procedia Computer Science*, 71, 2015, pp. 106-111.
- [60] Bryson, J.J. “Patiency is not a virtue: the design of intelligent systems and systems of ethics” *Ethics and Information Technology*, 20, ,2018, pp. 15-26.
- [61] Lin, P., Abney, K. and Bekey, G. “Robot ethics: Mapping the issues for a mechanized world” *Artificial Intelligence Review*, 175, 2011, pp. 942-949.