

# THE CHALLENGES OF A GLOBAL GOVERNANCE ON ARTIFICIAL INTELLIGENCE

Patricia Eliane da Rosa Sardeto<sup>1</sup> / Poliana Dequêch<sup>2</sup>

Patricia Eliane da Rosa Sardeto, PhD, University Professor at the Pontifical Catholic University of Paraná\_Câmpus Londrina (PUCPR), Jockey Club, 485 - Hípica, 86067-000 Londrina, Paraná, Brazil, patricia.sardeto@pucpr.br; <http://www.pucpr.br>

Poliana Dequêch, graduate student in ethics and bachelor's degree in law from the Pontifical Catholic University of Paraná - Câmpus Londrina, polianadequech@gmail.com, <http://www.pucpr.br>

**Keywords:** *Law, Artificial Intelligence, Education, Brazil, Global Governance*

**Abstract:** *Since Artificial Intelligence has the potential to impact human life in its entirety and globality, debates about global governance models of Artificial Intelligence are urgent. We already have some guidelines and general principles from(in) some countries and international entities, but so far, they cannot be considered a global governance proposal, in terms that UN proposed in 2023. In this way, our research aims to analyse how a global governance on AI can offer guaranties to be inclusive and collaborative, contributing so that everyone has a good experience with AI now and in the future. The problem addressed in this research is the interference in the process of education due to the use of new technologies, especially ChatGPT from the company Open AI. For this purpose, the systemic approach method proposed by Maturana and Varela is used, and the bibliographic method combined with comparative method is used as a procedural method. In way to achieve a global governance some initiatives are analysed and their challenges and contributions pointed out. The global governance on AI needs to come from different countries, entities, areas (tech/social/legal) as soon as possible because governance needs to learn to be fast, in order to follow the advancement of technology.*

## 1. Introduction

Artificial Intelligence has been the great tool used to delegate tasks that previously belonged to humans. Just as mechanization occurred in the field, it is natural for this process to intensify with the evolution of technologies.

But the delegation of human functions is not always healthy, safe, and among them the use of ChatGPT on education is discussed with some data and contributions.

There is a growing use of ChatGPT and it is necessary to understand the benefits and harms of this technology and the importance of global governance in the development of technologies that may interfere with the development of students.

The present work aims to understand the damage that Artificial Intelligence can cause to the development of students and consequently the need for global governance for the development of Artificial Intelligence, considering that its impact occurs collectively and not just in the place where the technology is developed.

## 2. The development of Artificial Intelligence and its potencial

Talking about Artificial Intelligence governance requires a multidisciplinary approach, as the term governance itself already demands multiple efforts and the term Artificial Intelligence is another one that, being so broad, needs to be narrowed down.

According to Coppin, Artificial Intelligence can be considered “the study of systems that act in a way that to any observer would appear to be intelligent”, noting that this definition does not cover the whole of AI as it

can deal with relatively simple problems or complex problems. Thus, the author also proposes another definition stating that Artificial Intelligence “involves using methods based on the intelligent behavior of humans and other animals to solve complex problems”.<sup>1</sup>

Russell and Norvig present four possible definitions of Artificial Intelligence, two in terms of fidelity to human performance, namely, systems that think like people and systems that act like people; two relating to an abstract and formal definition of intelligence, that is, systems that think rationally and systems that act rationally. Where a “rational agent is one who acts to achieve the best result or, when there is uncertainty, the best expected result”.<sup>2</sup>

Santaella adds that currently “these fundamentally rationalist objectives would need to be complemented by systems that feel like people, or better yet, simulate feeling like people”.<sup>3</sup>

In this sense, when Coppin distinguish between strong and weak Artificial Intelligence, highlights that for followers of strong Artificial Intelligence this would be precisely the objective, to create a robot with emotions and genuine consciousness, that is, a “computer that can literally think and be conscious in the same way that a human is conscious”.<sup>4</sup>

Since the 1950s the development of the Artificial Intelligence has gone through ups and downs, sometimes with memorable advances and sometimes with so-called “AI winter”. While advances in the development and application of Artificial Intelligence can and should be celebrated, we still find ourselves in human-level and weak AI. This means that Artificial Intelligence today already surpasses human intelligence in several domains, such as games, and although it may seem like a huge achievement, what we see in the game of chess can be applied to other games is that it’s about building a perfectly fine chess engine around a special-purpose algorithm. “When implemented on the fast processors that became available towards the end of the twentieth century, it produces very strong play. But an AI built like that is narrow. It plays chess; it can do no other”.<sup>5</sup>

Finally, Bostrom points out that recent years have seen a resurgence of interest in AI, leading to efforts towards artificial general intelligence<sup>6</sup>, considered by Nilsson to be strong AI, that variety of AI that attempts to mechanize human-level intelligence.<sup>7</sup> The fact is that today we have practically ideal conditions for its development, with super-powerful hardware, big data with thousands of pieces of information and increasingly real integration between knowledge, in addition to the development of sub-areas of AI such as computational neuroscience.

A big step was taken in November 2022, when OpenAI announced and made available for use by the general public, ChatGPT, a type of Generative AI, which, as its name suggests, creates/generates new content from inputs and outputs.

On the OpenAI website it is possible to access the document that presented ChatGPT. It is worth highlighting the description of the method used for its development:

We trained this model using Reinforcement Learning from Human Feedback (RLHF), using the same methods as InstructGPT, but with slight differences in the data collection setup. We trained an initial model using supervised fine-tuning: human AI trainers provided conversations in which they played both sides—the user and an AI assistant. We gave the trainers access to model-written suggestions to help them

---

<sup>1</sup> COPPIN, BEN. *Inteligência artificial*. Tradução e revisão técnica Jorge Duarte Pires Valério. Rio de Janeiro: LTC (2017), p. 4.

<sup>2</sup> RUSSELL, STUART J.; NORVIG, PETER. *Inteligência Artificial: uma abordagem moderna*. Trad. Daniel Vieira; Flávio Soares Corrêa da Silva. 4. Ed. Rio de Janeiro: LTC (2022).

<sup>3</sup> SANTAELLA, LÚCIA. *A inteligência artificial é inteligente?* São Paulo: edições 70 (2023).

<sup>4</sup> COPPIN, BEN. *Inteligência artificial*. Tradução e revisão técnica Jorge Duarte Pires Valério. Rio de Janeiro: LTC (2017), p. 5.

<sup>5</sup> BOSTROM, NICK. *Superintelligence: Paths, dangers, strategies*. Oxford University Press (2014), p. 14.

<sup>6</sup> It is not yet the proposal of “singularity” (Kurzweil) or machine superintelligence (Bostrom), both of which are at the level of super-human-level machine intelligence.

<sup>7</sup> BOSTROM, NICK. *Superintelligence: Paths, dangers, strategies*. Oxford University Press (2014), p. 22.

compose their responses. We mixed this new dialogue dataset with the InstructGPT dataset, which we transformed into a dialogue format.

To create a reward model for reinforcement learning, we needed to collect comparison data, which consisted of two or more model responses ranked by quality. To collect this data, we took conversations that AI trainers had with the chatbot. We randomly selected a model-written message, sampled several alternative completions, and had AI trainers rank them. Using these reward models, we can fine-tune the model using Proximal Policy Optimization. We performed several iterations of this process.<sup>8</sup>

ChatGPT “is trained to follow an instruction in a prompt and provide a detailed response”, still continues the presentation<sup>9</sup>. According to the technical definition of Nvidia, Generative AI enables users to quickly generate new content based on a variety of inputs. Inputs and outputs to these models can include text, images, sounds, animation, 3D models, or other types of data, that is why is possible to reinforce that ChatGPT is a tool in progress.<sup>10</sup>

Since its launch, ChatGPT is receiving upgrades. Its potential is extraordinary and has been scaring society in general and researchers and developers themselves, however we still do not talk about strong or general Artificial Intelligence.

### 3. The impact on education of new technologies in Brazil

As explained above, generative artificial intelligence has the potential to interfere more and more in human life. One of the points of interference identified in this research is the academic and intellectual development of young people.

ChatGPT emerged as an aid tool, but often serves as a crutch for cognitive deficits and presents new challenges to the regulatory world bringing to light the discussion about the ethical and responsible use of ChatGPT.

One of the main problems identified is the lack of accuracy and reliability of chatGPT and the increase in plagiarism among students. To Chung Kwan Lo “because ChatGPT is trained on a large corpus of data, it may be biased or contain inaccuracies”.<sup>11</sup>

Furthermore, the author raises other concerns: “Despite its success, ChatGPT has introduced new challenges and threats to education. With its ability to provide specific answers to user questions, it can be used to complete written assignments and examinations on behalf of students, leading to concerns about AI-assisted cheating”<sup>12</sup>

Simone Grassini explains that “the urgent need to address the impact of ChatGPT on the educational sector cannot be overstated, and the need for immediate action has been proposed. There is a pressing demand to adapt assessment practices and institutional protocols to manage the issues brought to the fore by the proliferation of AI-generated content in academic work”.<sup>13</sup>

In the article by Mustafizur Rahman and Yutaka Watanabe, the authors highlight several benefits and opportunities of using ChatGPT, but point out some threats that can be managed. In the words of the authors:

<sup>8</sup> OPEN AI. Introducing ChatGPT. Blog. <https://openai.com/blog/chatgpt>.

<sup>9</sup> OPEN AI. Introducing ChatGPT. Blog. <https://openai.com/blog/chatgpt>.

<sup>10</sup> NVIDIA. Generative AI. <https://www.nvidia.com/en-us/glossary/data-science/generative-ai/>.

<sup>11</sup> LO, CHUNG KWAN. „What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature“ *Education Sciences* 13, no. 4: 410. <https://doi.org/10.3390/educsci13040410> (2023).

<sup>12</sup> LO, CHUNG KWAN. „What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature“ *Education Sciences* 13, no. 4: 410. <https://doi.org/10.3390/educsci13040410> (2023).

<sup>13</sup> GRASSINI, SIMONE. Shaping the Future of Education: Exploring the Potential and Consequences of AI and ChatGPT in Educational Settings. *Education Sciences* 13, no. 7: 692. <https://doi.org/10.3390/educsci13070692> (2023).

However, although ChatGPT is a powerful tool that can generate impressive responses on a variety of topics, it still has certain limitations, such as a lack of common sense, potential bias, difficulty with complex reasoning, and inability to process visual information. It is important to keep in mind the limitations of ChatGPT when using it, and it should not be relied upon blindly. In addition, the ethical implications (e.g., bias and discrimination, privacy and security, misuse of technology, accountability, transparency, and social impact) of ChatGPT are complex and multifaceted and should be carefully considered. Despite the various difficulties and challenges, we believe that the risks discussed can be effectively managed and must be addressed to provide reliable and equitable access to LLMs for educational and research purposes.<sup>14</sup>

In a survey carried out in March 2023 by the Walton Foundation in the United States among students aged 12 to 17 and their teachers in these groups, it was found that teachers are using ChatGPT more than students, at a rate of 51% to 31%. And this is because they understand that new technology can be an important tool in the teaching and learning process, functioning as an “individual tutor” for students and helping those with more difficulty, especially groups of students who are more vulnerable and more impacted by the consequences of covid 19 pandemic. Additionally, ChatGPT can help plan lessons and provide creative lesson insights.<sup>15</sup>

Analyzing this research, Ronaldo Lemos compares it with the Brazilian reality and suggests that the reason why Brazilian teachers are quickly adopting AI is because they realize that the technology can save time in their daily lives, since they are overwhelmed with work. AI can help in developing lesson plans, organizing curricula, summarizing and organizing complex texts. Therefore, he concludes that the trend is that in Brazil AI, and in this case ChatGPT, is also adopted faster by teachers than by students.<sup>16</sup>

From the perspective of an underdeveloped country, such as Brazil, it is understood that the management of the risks of using ChatGPT must be done through global governance from some body that is capable of inducing international cooperation, as is the case with the United Nations.

In a lecture at the University of São Paulo, Brazil, rector Carlos Gilberto Carlotti Junior argued that legislative bodies play a crucial role in the transparency of artificial intelligence systems. For him, legislators will have to define some standards to protect society, as “a system cannot be so closed that it is not known where the solution it presents comes from”. The biggest challenge will be methodological, according to him. And concluded “We need to understand the tools and learn how to use them. We need to know what information we should ask for and how to ask for it.”<sup>17</sup>

Considering the legislative differences between each country and the global impact of technologies, the importance of establishing global governance becomes important.

#### **4. The need for a local and global governance**

There are many problems related to the introduction of technology tools into everyday education. As mentioned in this work, ethics, reliability, accuracy, cheating, plagiarism, among other problems. For these reasons, we focus on proposing global governance, as the impacts are global, but a local governance as well, because specific issues need to be considered.

---

<sup>14</sup> RAHMAN, MD. MOSTAFIZER, and YUTAKA WATANOBÉ. „ChatGPT for Education and Research: Opportunities, Threats, and Strategies“ *Applied Sciences* 13, no. 9: 5783. <https://doi.org/10.3390/app13095783> (2023).

<sup>15</sup> *Foundation, Walton Family*. ChatGPT used by teachers more than students, new survey from Walton Family Foundation finds. <https://www.waltonfamilyfoundation.org/chatgpt-used-by-teachers-more-than-students-new-survey-from-walton-family-foundation-finds>.

<sup>16</sup> LEMOS, RONALDO. Professores usam IA mais que os alunos. ITS. <https://itsrio.org/pt/artigos/professores-usam-ia-mais-que-os-alunos/>.

<sup>17</sup> BELESSA, MAURO. Os desafios do ChatGPT ao ensino e à pesquisa. <http://www.iea.usp.br/noticias/os-desafios-do-chatgpt-para-a-universidade?searchterm=Chat+GPT> (2023).

We already have some guidelines and general principles from(in) some countries and international entities, but so far, they cannot be considered a global governance proposal, in terms that UN proposed in 2023.

Just for example, we can point out some regulation/guideline initiatives on AI, including those by states as well as regional and intergovernmental processes such as the EU, the G7, the G20, UNESCO, and the OECD, among others. Especially since 2019, the year that effective appears several contributions, the G7 presented its Global Partnership on AI, the EU its Ethics Guidelines for Trustworthy/Policy and Investment Recommendations for Trustworthy AI, the OECD its Recommendation on AI, the G20 its AI Principles, the WEF its AI Government Procurement Guidelines.

In 2020 the EU presented its White Paper on AI and in April 2021, the European Commission proposed the first EU regulatory framework for AI. It says that AI systems that can be used in different applications are analysed and classified according to the risk they pose to users. The different risk levels will mean more or less regulation. Recently, on December 9 2023, Parliament reached a provisional agreement with the Council on the AI act. The agreed text will now have to be formally adopted by both Parliament and Council to become EU law.

Sure that the EU's effort to approve the first AI law contributes a lot in a way to achieve a such expected global governance.

However, the UN proposal for global governance appears to create a more comprehensive and diverse framework. Initially proposed in 2020 as part of the Secretary-General's Roadmap for Digital Cooperation (A/74/821), the multi-stakeholder High-level Advisory Body on Artificial Intelligence was formed in October 2023 to undertake analysis and advance recommendations for the international governance of AI.

Its mission is not only answer how to govern AI today, but also how to prepare governance institutions for an environment in which the pace of change is only going to increase. Following the roadmap planned by the High-level Advisory Body on AI for 2023 (Aug 2023 – call for experts; Oct 2023 AI advisory body formed; nov 2023 initial consultations; end 2023 – interm report released), we can check the main points highlighted in the interm report, published on december 2023.

First of all, the importance of the global governance, pointed out in item 39:

There is, today, no shortage of guides, frameworks, and principles on AI governance. Documents have been drafted by the private sector and civil society, as well as by national, regional, and multilateral bodies, with varying degrees of impact. In technology terms, governance efforts have been focused on data, models, and benchmarks or evaluations. Applications have also been under focus, especially where there are existing sectoral governance arrangements, say for health or dual-use technologies. These efforts can be anchored in specific governance arrangements, such as the EU AI Act or the U.S. Executive Order and they can be associated with incentives for participation and compliance.<sup>18</sup>

For now we have the preliminary recommendations offered in this interm report focus on the principles that should guide the formation of new global governance institutions for AI and the broad functions such institutions would need to perform.

To know them:

#### Guiding Principles

1. AI should be governed inclusively, by and for the benefit of all
2. AI must be governed in the public interest
3. AI governance should be built in step with data governance and the promotion of data commons

---

<sup>18</sup> *United Nations*. Interm report: governing AI for Humanity. [https://www.un.org/sites/un2.un.org/files/ai\\_advisory\\_body\\_interim\\_report.pdf](https://www.un.org/sites/un2.un.org/files/ai_advisory_body_interim_report.pdf).

4. AI governance must be universal, networked and rooted in adaptive multistakeholder collaboration
5. AI governance should be anchored in the UN Charter, International Human Rights Law, and other agreed international commitments such as the Sustainable Development Goals

#### Institutional Functions

1. Assess regularly the future directions and implications of AI
2. Reinforce interoperability of governance efforts emerging around the world and their grounding in international norms through a Global AI Governance Framework endorsed in a universal setting (UN)
3. Develop and harmonize standards, safety, and risk management frameworks
4. Facilitate development, deployment, and use of AI for economic and societal benefit through international multi-stakeholder cooperation
5. Promote international collaboration on talent development, access to compute infrastructure, building of diverse high-quality datasets, responsible sharing of opensource models, and AI-enabled public goods for the SDGs
6. Monitor risks, report incidents, coordinate emergency response
7. Compliance and accountability based on norms.<sup>19</sup>

Until middle 2024 the final report should be finished and presented in September 2024 in the Summit of the Future, in order to the UN Member States consider a Global Digital Compact.

The efforts of the states as well as regional and intergovernmental entities should be recognized, but we believe that the theme “education” is too important to be left for later. In case of the Interim report, this was what happened. The education is the basis of our society and the impacts of the Artificial Intelligence appear first on children and young people, the digital native generation.

The discussion and possible guidelines/solutions for de AI need to cover the generation of the future, that is why the education needs to be prioritized. In this sense, for establishing this global governance we propose the establishment of universal ethical principles related to education, global standards to be incorporated by the public policies of the countries, assessment of the impact on various sectors of civil society, independent monitoring and periodic review with feedback.

Locally, in case of the poor and emerging countries that are taking place in the development of this new technology, but as passive “players“, because in most times their population is being used to “feed” new applications on AI, a contribution more effective and opportunities to participate with voice and power of decision. Other issues such as multilateral engagement, cooperation and connection with personal data protection regulations, ethical framework and new human rights need to be analyzed as well.

It is believed that with the establishment of global governance, the development of technologies will be carried out in the most beneficial way possible, without negatively impacting countries that are not protagonists in technological development.

## 5. Conclusion

In this research we concluded that new technology tools have some consequences that must be evaluated and managed. The proposed form of management is bases on global governance, considering the global impacts of technology.

---

<sup>19</sup> *United Nations*. Interim report: governing AI for Humanity. [https://www.un.org/sites/un2.un.org/files/ai\\_advisory\\_body\\_interim\\_report.pdf](https://www.un.org/sites/un2.un.org/files/ai_advisory_body_interim_report.pdf).

The challenges are many, especially on education. There are specific problems and difficulties, according to each country and culture, but issues related to bias, ethics, reliability, accuracy, cheating, plagiarism are global and need to face globally.

In relation to the United Nations, the creation of guidelines and regulations is of utmost importance especially because they have a diverse and global vision, considering their relevance and strength in the most diverse countries.

## 6. References

- BELESSA, MAURO. Os desafios do ChatGPT ao ensino e à pesquisa. <http://www.ica.usp.br/noticias/os-desafios-do-chatgpt-para-a-universidade?searchterm=Chat+GPT> (2023)
- BOSTROM, NICK. *Superintelligence: Paths, dangers, strategies*. Oxford University Press (2014)
- COPPIN, BEN. *Inteligência artificial. Tradução e revisão técnica Jorge Duarte Pires Valério*. Rio de Janeiro: LTC (2017)
- Walton Family Foundation*. ChatGPT used by teachers more than students, new survey from Walton Family Foundation finds. <https://www.waltonfamilyfoundation.org/chatgpt-used-by-teachers-more-than-students-new-survey-from-walton-family-foundation-finds>.
- GRASSINI, SIMONE. Shaping the Future of Education: Exploring the Potential and Consequences of AI and ChatGPT in Educational Settings. *Education Sciences* 13, no. 7: 692. <https://doi.org/10.3390/educsci13070692> (2023)
- LEMOS, RONALDO. Professores usam IA mais que os alunos. ITS. <https://itsrio.org/pt/artigos/professores-usam-ia-mais-que-os-alunos/>.
- LO, CHUNG KWAN. „What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature“ *Education Sciences* 13, no. 4: 410. <https://doi.org/10.3390/educsci13040410> (2023)
- NVIDIA. Generative AI. <https://www.nvidia.com/en-us/glossary/data-science/generative-ai/>.
- OPEN AI. Introducing ChatGPT. Blog. <https://openai.com/blog/chatgpt>.
- RAHMAN, MD. MOSTAFIZER/YUTAKA WATANOBE. „ChatGPT for Education and Research: Opportunities, Threats, and Strategies“ *Applied Sciences* 13, no. 9: 5783. <https://doi.org/10.3390/app13095783> (2023)
- RUSSELL, STUART J./NORVIG, PETER. *Inteligência Artificial: uma abordagem moderna*. Trad. Daniel Vieira; Flávio Soares Corrêa da Silva. 4. Ed. Rio de Janeiro: LTC (2022)
- SANTAELLA, LÚCIA. *A inteligência artificial é inteligente?* São Paulo: edições 70 (2023)
- SCHWAB, KLAUS. *The Fourth Industrial Revolution*. New York: Currency (2017)
- United Nations*. Interim report: governing AI for Humanity. [https://www.un.org/sites/un2.un.org/files/ai\\_advisory\\_body\\_interim\\_report.pdf](https://www.un.org/sites/un2.un.org/files/ai_advisory_body_interim_report.pdf)

