

## AI Governance Study and Proposal

2023 September

Doohee You, PhD

Apple Co LTD, Data Science/ AI Manager

(Report is not representing company's opinion and fully accredited to the author)

# Introduction

Artificial intelligence (AI) is rapidly transforming our world, with the potential to improve our lives in many ways. However, AI also raises a number of important governance challenges, such as:

- How can we ensure that AI is used for good and not for harm?
- How can we prevent AI from being used to discriminate or violate human rights?
- How can we ensure that AI is developed and used in a transparent and accountable way?

These challenges are complex and there are no easy answers. However, it is clear that we need to develop a global AI governance framework to address them.

## 1. The Need for Global AI Governance

The challenges of AI are not limited to any one country or region. They are global in nature and require a global response. A global AI governance framework would allow us to share knowledge and best practices, coordinate our efforts, and build trust.

There are a number of different approaches to global AI governance. One approach is to develop a binding international treaty. This would be the most comprehensive and effective way to address the challenges of AI. However, it is also the most difficult to achieve, as it would require the agreement of all countries.

Another approach is to develop a set of voluntary guidelines or principles. This would be a less ambitious goal, but it would still be a valuable step forward. Voluntary guidelines could provide a common framework for countries to develop their own AI policies.

### I. The Role of Multilateral Organizations

Multilateral organizations can play an important role in developing a global AI governance framework. These organizations have the expertise and resources to facilitate international cooperation and build consensus.

The OECD has already developed a set of AI Principles that can serve as a foundation for global AI governance. These principles focus on issues such as human-centered AI, fairness and non-discrimination, transparency and accountability, and responsibility.

The WHO has also developed guidance on data and AI governance in the health sector. This guidance focuses on issues such as privacy and security, data sharing, and the development of ethical AI tools.

## **II. AI Governance Dilemmas**

There are a number of dilemmas that need to be addressed in the development of a global AI governance framework. These dilemmas include:

- The tension between innovation and safety. AI is a rapidly developing field, and it is important to strike a balance between encouraging innovation and ensuring that AI is used safely.
- The tension between privacy and transparency. AI systems often collect and use large amounts of data, which raises concerns about privacy. However, transparency is also important to ensure that AI systems are accountable and that their decisions can be understood.
- The tension between individual rights and collective interests. AI systems can have a significant impact on individuals, but they can also be used to promote the common good. It is important to find ways to balance these competing interests.

## **III. Regulation of Large Language Models**

Large language models (LLMs) are a type of AI that are capable of generating human-quality text. They have the potential to be used for a wide range of purposes, including generating creative content, translating languages, and answering questions.

However, LLMs also raise a number of potential risks, such as the spread of misinformation and the creation of deepfakes. It is therefore important to consider how to regulate LLMs in a way that balances the risks and benefits.

One possible approach is to use a sandbox technique. A sandbox is a controlled environment where new technologies can be tested and evaluated before they are released to the public. This approach could be used to test LLMs to ensure that they are not being used for harmful purposes.

## **IV. The Use of Sandboxes in AI Governance**

Sandboxes can be used to address a number of challenges in AI governance. They can be used to:

- Test new AI technologies in a controlled environment
- Identify and mitigate potential risks
- Build public trust in AI
- Facilitate international cooperation

Sandboxes can be implemented at different levels, from the national level to the international level. They can be used by governments, businesses, and civil society organizations.

## 2. The Future of Global AI Governance

The future of global AI governance is uncertain. However, it is clear that we need to develop a framework that is comprehensive, effective, and adaptable to the rapid pace of technological change.

Multilateral organizations can play an important role in developing and implementing a global AI governance framework.

### I. Bilateral Initiatives

There are a number of different initiatives underway to promote global AI governance. These initiatives are taking place at the bilateral, multilateral, and inter-regional levels.

Some of the most active bilateral initiatives on AI governance are taking place between the United States and China. These two countries are the leading developers of AI technology, and they have a shared interest in ensuring that AI is used responsibly.

In 2019, the United States and China issued a joint statement on AI principles. The statement outlined a number of common principles for the development and use of AI, such as the need for transparency, accountability, and non-discrimination.

The United States and China are also working together on a number of other initiatives on AI governance, such as the development of common standards for data privacy and security.

### II. Multilateral Initiatives

There are a number of multilateral initiatives on AI governance underway. One of the most important of these initiatives is the OECD's AI Principles. The OECD Principles are a set of voluntary guidelines for the development and use of AI. They focus on issues such as human-centered AI, fairness and non-discrimination, transparency and accountability, and responsibility.

The OECD Principles have been endorsed by over 50 countries, and they are increasingly being used as a reference point for national AI policies.

Another important multilateral initiative is the UN's AI for Good Global Summit. The Summit is a forum for governments, businesses, and civil society to discuss the challenges and opportunities of AI. The Summit has been held twice, in 2019 and 2021.

### III. Inter-regional Initiatives

There are also a number of inter-regional initiatives on AI governance underway. One of the most important of these initiatives is the AI for Humanity Partnership. The Partnership is a collaboration between the European Union, the United States, and Japan.

The AI for Humanity Partnership aims to promote the responsible development and use of AI for the benefit of humanity. The Partnership has developed a set of principles for AI governance, and it is working to implement these principles in its own policies and practices.

#### **IV. Philosophical Approaches**

There are a number of different philosophical approaches to AI governance. One approach is to focus on the technical aspects of AI, such as the development of standards and regulations. Another approach is to focus on the ethical implications of AI, such as the potential for bias and discrimination.

A third approach is to focus on the social and political implications of AI, such as the impact of AI on jobs and the economy.

There is no single "correct" philosophical approach to AI governance. The best approach will likely vary depending on the specific context.

#### **V. Critical Perspectives**

There are a number of critical perspectives on AI governance. Some critics argue that AI governance is impossible, as AI is too complex and unpredictable. Others argue that AI governance is unnecessary, as the market will self-regulate.

Still others argue that AI governance is necessary, but that it should be done in a way that does not stifle innovation.

## **3. Recommendations**

There are a number of recommendations for global AI governance. Some recommendations focus on the need for international cooperation. Others focus on the need for transparency and accountability.

Still others focus on the need to address the ethical implications of AI, such as the potential for bias and discrimination.

The best way to promote global AI governance is likely to involve a combination of different approaches. It is also important to involve a wide range of stakeholders in the development of AI governance frameworks. This will help to ensure that these frameworks are legitimate and effective.

## **4. Existing Models**

There are a number of different models for global AI governance that have been proposed. Some of these models are based on existing international law, while others are more innovative and forward-looking.

One of the most common models for global AI governance is to build on existing international law. This approach has the advantage of being familiar and well-established, but it may not be sufficient to address the unique challenges posed by AI.

For example, the UN Convention on Cybercrime (2001) addresses some of the legal issues related to AI, such as data protection and cybersecurity. However, the Convention does not specifically address the ethical implications of AI, such as the potential for bias and discrimination.

Another approach to global AI governance is to develop new international law. This approach would allow for the development of more specific and tailored rules that address the challenges of AI. However, it is also the most difficult and time-consuming approach.

## **I. Innovative Models**

In addition to existing and new international law, there are a number of innovative models for global AI governance that have been proposed. These models include:

- The development of a global AI governance treaty: This would be a binding agreement between countries that sets out the rules and principles for the development and use of AI.
- The creation of a global AI governance agency: This would be an independent body that would oversee the development and use of AI.
- The use of sandboxes: This would involve allowing new AI technologies to be tested in a controlled environment before they are released to the public.
- The development of ethical guidelines for AI: This would involve developing a set of principles that would guide the development and use of AI in a responsible and ethical way.

## **II. Surveys and Analyses**

There have been a number of surveys and analyses of different proposals for global AI governance. These studies have identified a number of challenges and opportunities associated with each approach.

One of the main challenges is the need to balance the different interests involved in AI governance. These interests include the interests of governments, businesses, civil society, and individuals.

Another challenge is the need to ensure that global AI governance is adaptable to the rapid pace of technological change. AI is a rapidly evolving field, and it is important to develop frameworks that can keep up with the latest developments.

# **5. Conclusion**

The development of global AI governance is a complex and challenging task. However, it is essential if we want to ensure that AI is used for good and not for harm. There are a number of different models for global AI governance that have been proposed, and it is important to continue to explore and develop these models.

The best approach to global AI governance will likely involve a combination of different approaches. It is also important to involve a wide range of stakeholders in the development of global AI governance frameworks. This will help to ensure that these frameworks are legitimate and effective.