

## Club de Madrid/Boston Global Forum

### POLICY LAB

## FUNDAMENTAL RIGHTS IN AI & DIGITAL SOCIETIES: TOWARDS AN INTERNATIONAL ACCORD

### Sub-committee 3

The elements & process for an international legal framework to protect  
fundamental rights in AI & digital spaces.

### Issues Paper for the workshop on 7-9 September 2021

- What problem are we trying to address?
- Why law? Why AI ethics is not enough
- Why domestic law is not sufficient and why we need a global agreement
- Consensus is possible
- Steps towards a global multilateral Agreement on AI Governance with the aim to ensure governability of States and a stable international order, collective rights and self-determination of humans and their fundamental rights in the age of AI
  - A mandate to explore the opportunity of a global agreement
  - A process of exchange of information on the subject matter, building inter alia on existing processes within UN Agencies
  - Principles for the process of work towards a global agreement
  - Principles on how to delimit the content of the agreement
    - The aim to maintain governability and related questions of control
    - The aim to maintain self-determination of humans
    - The aim to protect global Human Rights
    - A focus on the specific technological risks resulting from the use of self-learning Algorithms and Data
  - Institutional arrangements
    - Mechanisms to create mutual trust
    - Mechanisms of reporting
    - Mechanisms of decision making
    - Mechanisms of dispute settlement and enforcement
  - Relation to pre-existing international law
  - Role of the United Nations Security Council
  - Existing International Law to be inspired by:
    - United nations convention on the law of the sea (UNCLOS)
    - Treaties concluded under the auspices of the International Atomic Energy Agency (IAEA) or related to its work such as those on nuclear safety, liability, non-proliferation
  - Sources of Inspiration in terms of ongoing work on legally binding AI governance:
    - Work in the Council of Europe on a Convention on AI

- Work in the European Union on an AI Act
- Chinese work on AI Regulation

### **What problem are we trying to address?**

The opportunities and risks of Artificial Intelligence (AI) seem limitless. While some believe that many of the most challenging problems of this world can be solved through or with the help of AI, others, such as Bill Gates, consider this technology also a risk alike to atomic power and atomic weapons.

There are a number of reasons why AI poses a global risk: First, AI is analysing data through Algorithms, which are supposed to learn, and thus improve their performance, beyond capabilities of humans. While today such performance of AI beyond capabilities of humans is normally limited to very specific tasks, there is a trend to the broadening of these tasks. It cannot be excluded anymore that a general AI is being developed, which surpasses all aspects of human intelligence and eventually thus could dominate humankind in all respects. The problem of control of AI in order to ensure that humans do not become objects of machine control and the ability of states to govern is not undermined has already been amply described in science. It is clear that technology alone cannot deliver such control to a sufficient extent.<sup>1</sup>

Second, much of Artificial Intelligence is and will be delivered and deployed via the Internet, across the borders of this world. It is thus a technology crossing borders in the virtual space without effective controls.

Third, AI is a technology, which is being developed by global corporations and states for various purposes and not limited to a sector. It is a multi-purpose technology, which has potential to scale globally in most areas of our lives, ranging from education and health via the production and delivery of media content and opinions, important for democracy, right through to the management of all essential infrastructures and military purposes. It is the sensitivity of the multiple contexts within which AI is deployed which requires to give its good functioning and governance highest attention. It is also clear that those who control the functioning of AI in sensitive sectors, including both states and global corporations, will command greatest power in this world.

Fourth, AI can be embedded in autonomous machines, which may cross borders in trade, crime or military operations, thus combining classical physical safety and security risks with the new risks of AI.

Fifth, AI performs often nothing more than large scale and rigorous optimisation. The large scale and global reach of AI programmes for optimisation may create major problems if they create disadvantages for many which are often invisible. For example, if AI optimises for one group, namely the owners of a cooperation, it may at the same time create major detriments to another group, for example workers. The power and scale of such rigorous, and often not visible optimisation, is a key global risk of AI, as it can create huge detriment to large groups of people on an unprecedented scale.

It is important to consider this cross border, global and multipurpose nature of AI in any attempt to assess risks and opportunities of AI. The world has become a community of risk, not only relating to COVID, but also relating to AI. And the scarcest resource is not the one next great idea in terms of a technological solution to a major global challenge, but the scarcest resource relating to AI is the ability

---

<sup>1</sup> Iyad Rahwan a.o., Superintelligence Cannot be Contained: Lessons from Computability Theory, Journal of Artificial Intelligence Research 70 (2021) 65-76, [https://pure.mpg.de/rest/items/item\\_3020363\\_8/component/file\\_3283858/content](https://pure.mpg.de/rest/items/item_3020363_8/component/file_3283858/content); see also Stuart Russel, Human Compatible – AI and the problem of control, 2019, with references to reviews and different language versions at <https://people.eecs.berkeley.edu/~russell/hc.html>.

to agree, both within states and among states, on how to govern this new powerful and globally scaled multipurpose technology to the benefit of states and mankind.

It is before this background that we are discussing how to give an impetus to the international community to start work on a global multilateral agreement on the governance of AI.

In a world where even Member States of the UN opt for “club”/plurilateral models such as the G-20, and taking account of the reality that only a few countries in the world are actually producing AI products, it will be important to identify the right forum for such work to start. The fact that all states and all people of this world are likely to be affected by an omnipresent, multipurpose technology like AI is an argument to be considered in favour of placing such initial work under the auspices of the UN. In the course of this work, the question of the right forum to make progress will certainly arise repeatedly. One may also ask whether there is a group of countries which could act as catalysts for a global agreement.

Geo-techno politics must be factored in. This includes the growing discontent with China’s growing role in UN agencies and on AI innovations and their use to control people and society.

Also, it is generally considered that developing nations align with China’s critiques of the prevailing approach to global governance/multilateral governance. The ‘G77’ is often overlooked in these discussions. While not necessarily major AI producers, they are the market base for which the major producers (US, Europe, China) will be gunning for market share. As they are thus affected, their voice must be heard.

### **Why law? Why AI ethics is not enough**

How a rights based global agreement will fare in the world that is laden with norms, non binding agreements and models of AI Governance outside the scope of state law is a crucial question. There have been over the last decade numerous attempts to create an Ethics of Artificial Intelligence. Corporations developing AI, in order to shoulder their responsibilities and to give orientation to their Engineers, but also states and multilateral organisations, either to prepare or to substitute legislation, have driven these efforts. The number of publications on AI Ethics in the academic field has become hard to follow. There are now more than 80 catalogues of Ethics for AI. Many of their principles overlap in general, but there are also many differences in important details.

Some of the Ethics Codes of Conduct have been accompanied by governance structures within corporations. However, some of these structures have been abandoned again or have been criticized as ineffective. There is also an emerging concern among business about free riders who will not commit to voluntary codes of Ethics, thus putting into question the level playing field of markets and competition.

The ethics community has a moral hazard when it comes to advocating for law as an instrument to govern AI. It is notable however on the other hand that some ethics committees have called for law to be put in place, in order to ensure the democratic legitimacy, the binding nature of the rules and enforceability of AI governance rules against even the most powerful corporations but also the many potential free riders.

Considering the high potentials as well as high risks associated with AI, and the concerns of business relating to a level playing field for AI, in the European Union a consensus has emerged that binding law is necessary, and that the previous AI ethics and governance exercises on the level of the European Union and elsewhere were a good preparation in terms of identifying the challenges which need to be addressed in law. The European Commission has for this reason proposed an AI Act, which is

presently being negotiated by the legislators. Similar considerations have led the broader Council of Europe to start exploratory work towards a binding Convention on AI.

Much of the private self-governance and non binding rules have value in terms of issue spotting and directional orientation. However, these non-binding instruments alone have not delivered the legal certainty and level playing field, which are necessary to both ensure that technology development can thrive and that this progress is to the benefit of humankind, thus in particular that risks arising from the new technologies and related business models are sufficiently addressed and mitigated.

### **Why domestic law is not sufficient and why we need a global agreement**

On the global Level, both the OECD and the G-20 have already recognised the need for common principles on AI Governance. However, the texts adopted in these fora, while giving orientation to corporations, engineers and domestic lawmakers, suffer from the same deficiency as ethics codes: They only have a character of a political appeal, do not carry democratic legitimacy nor binding nature or enforceability. Executives, not legislators, which normally ratify international law obligations, have thus signed them.

Domestic law, which only the European Union is presently negotiating among legislators, will not be able to fully address the global scalability and cross border nature of AI. In order for all states to be certain that the ability to govern will not be undermined through AI being either used with intent for this purpose or getting out of control and thus undermining governability, it is necessary to create rules and structures through which states can support each other in maintaining control over AI, to the benefit of governability and human rights, thus mankind. Only international law which sets out basic substantial principles for this purpose as well as institutions and mechanisms sufficiently developed to be able to deal with the power accruing to those developing and controlling AI will be able to serve this purpose.

It is important that the great powers of this world as well as small states all sign up to such a global agreement, as AI can be developed and deployed all over the world, with impacts in all other parts of this world. Legitimacy for establishing an international legal framework arises out of the common interest in governability of states, a peaceful international order and giving effect to existing rules of international law in the technical age, in substantive terms, and in procedural terms, as in all international law, from the ratification by legislators.

Considering that much power arising with AI is in the hands of private companies, the question arises whether these companies should be directly bound by a global agreement and how this could be made possible. Rules directly applicable to private parties can arise from international law. The alternative to this is an international agreement in which states take the obligation to enforce certain rules against private actors under their jurisdiction. In *fine* the state enforcement against private parties of any rules agreed will be key in both cases to give effect to binding rules. <sup>2</sup>

### **Consensus is possible**

Since 2019, successive exercises of consensus building on principles for AI have demonstrated that a global consensus is possible. While none of these texts is binding, they show that the international community is learning about the opportunities and risks of AI. And they express a need to increase

---

2

[https://scholarship.law.georgetown.edu/cgi/viewcontent.cgi?article=1987&context=facpub&httpsredir=1&ref](https://scholarship.law.georgetown.edu/cgi/viewcontent.cgi?article=1987&context=facpub&httpsredir=1&referer=)

precaution, in light of the COVID experience as well as increasing potentialities of AI and related risks studies, while keeping the way open for the development of AI in the best interest of mankind. In 2019 the OECD Council adopted Recommendations on Artificial Intelligence<sup>3</sup>, after long negotiations. While the US had initially opposed the adoption of the principles, it eventually agreed, in a remarkable turn of position. On this basis, the G 20 adopted its human – centred AI principles.<sup>4</sup>

In the run up to the EU – US Summit in Brussels on 14 June, a group of academics from the US and the EU produced a Manifesto “In defence of Democracy and the Rule of Law in the age of AI”. This manifesto sets out proposals for legal action related to AI and digital technology and thus marks a new emerging consensus across the Atlantic.<sup>5</sup> Academic consensus does not often lead to political consensus. But the manifesto, signed in the meantime by academics also from other continents, is a sign that the unbridgeable gap which existed in the time of President Obama on regulating the digital economy between the EU and the US is being closed. The manifesto followed the invitation to the US by President of the European Commission Ursula Von der Leyen to start work on an AI Agreement: “We want to set a blueprint for regional and global standards aligned with our values: Human rights, and pluralism, inclusion and the protection of privacy”, she said when accepting the World Leader for global Peace and Security award from Governor Dukakis at the Boston Global Forum.<sup>6</sup>

More recently, in July 2021, the UNESCO Representatives of Member States agreed on a draft recommendation on AI governance, to be submitted to the General Conference of UNESCO Member States in November 2021 for adoption. UNESCO is now calling for an international regulatory framework to ensure that AI benefits humanity as a whole.<sup>7</sup>

Both China and Russia are Members of UNESCO. It remains to be seen whether on ministerial level they join the UNESCO consensus. There are clear however signs of the Chinese government increasingly understanding the need to regulate the digital economy, for various reasons, which include governability in light of the technological and economical power being concentrated through AI. Also the Russian Prime Minister Putin has taken a stance on the need for rules and limitations on AI.<sup>8</sup> While these are just initial indications, maintaining governability in light of technologies, which may become autonomous to an extent that makes control impossible, may be a common concern of many governments, including China and Russia. Maintaining governability could thus be a starting point for work towards a consensus on a broader rights based global agreement on binding rules for AI.

### **Next Steps and issues to be addressed on the way to a global agreement on AI**

- A mandate to explore the opportunity of a global agreement

This mandate should be focussed on making more concrete the understanding of states how far they are already becoming a global community of risks, and thus how far their common interest in global and binding governance rules as well as international mechanisms to enforce these rules in practice go.

- A process of exchange of information on the subject matter

---

<sup>3</sup> <https://www.oecd.org/going-digital/ai/principles/>

<sup>4</sup> <https://www.mofa.go.jp/files/000486596.pdf>, point 3.

<sup>5</sup> <https://www.aiathens.org/manifesto>

<sup>6</sup> [https://ec.europa.eu/commission/presscorner/detail/en/speech\\_20\\_2402](https://ec.europa.eu/commission/presscorner/detail/en/speech_20_2402)

<sup>7</sup> <https://en.unesco.org/artificial-intelligence/ethics>

<sup>8</sup> <https://voicebot.ai/2020/12/10/russian-president-vladimir-putin-rejects-idea-of-ai-politician-in-interview-with-sberbank-voice-assistant/>

The process of exchange of information should go beyond the United Nations SG Roadmap for Digital Cooperation,<sup>9</sup> in which AI is one topic. It should focus on existing and future capabilities of AI and the related risks, in particular the problem of control of AI and the challenge of establishing global governance of AI. The process should be open to states, academia, civil society and business. Its aim is to increase the global understanding of risks related to AI and what global governance rules and movements to implement such rules (e.g. a Global Alliance for Digital Governance) will be necessary.<sup>10</sup>

- Principles for the process of work towards a global agreement

The work towards a global agreement should be open to all interested stakeholders, but run by the Secretary General and a group of lead States. It should in a first step aim to identify existing principles under international law and international human rights law which may be put in question or suffer in their implementation from AI and what legal rules and mechanisms are necessary to address these challenges.

- An early agreement should be sought on principles on how to delimit the content of the agreement
  - The aim to maintain governability and related questions of control of AI by humans (rather than humans being controlled by AI).
  - The aim to maintain self-determination of humans
  - The aim to protect universal Human Rights aligned with the HR conventions
  - A focus on the specific technological risks resulting from the use of self-learning Algorithms and Data, already identified by professional associations<sup>11</sup> and science<sup>12</sup>, as well as a complementary, lateral risk assessment, incorporating (geo)political, economic and sociocultural risks posed by AI.
- An early agreement should be sought that Institutional arrangements must be put in place to ensure compliance with the legal principles of the agreement, such as:
  - Mechanisms to create mutual trust
  - Mechanisms of reporting
  - Mechanisms of decision making among parties
  - Mechanisms of dispute settlement and enforcement
  - Mechanisms to enforce directly against private parties under certain conditions, given their relevance in the field of AI

The Relation to pre-existing international law and a possible role of the United Nations Security Council in relation to the enforcement of the AI Agreement should be explored.

<sup>9</sup> <https://www.un.org/en/content/digital-cooperation-roadmap/>

<sup>10</sup> See on this also **the United Nations Centennial Initiative and the volume of reports on “Remaking the World – Toward an Age of Global Enlightenment”**, 15. July 2021, <https://bostonglobalforum.org/publications/remaking-the-world-the-age-of-global-enlightenment-2/>.

<sup>11</sup> <https://standards.ieee.org/industry-connections/ec/autonomous-systems.html>

<sup>12</sup> [https://futureoflife.org/data/documents/research\\_priorities.pdf?x72900](https://futureoflife.org/data/documents/research_priorities.pdf?x72900)

- Existing International Law to be inspired by in terms of the challenges faced and the substantial as well as institutional solutions found are, among others:
  - United nations convention on the law of the sea (UNCLOS)
  - Treaties concluded under the auspices of the International Atomic Energy Agency (IAEA) or related to its work such as those on nuclear safety, liability, non-proliferation.
  
- Sources of Inspiration in terms of ongoing work on legally binding AI governance:
  - Work in the Council of Europe on a Convention on AI<sup>13</sup>
  - Work in the European Union on an AI Act<sup>14</sup>
  - Chinese work on AI Regulation<sup>15</sup>

Paul Nemitz, Brussels 1.09. 2021

The author here expresses his personal opinion and not necessarily that of the European Commission.

---

<sup>13</sup> <https://www.coe.int/en/web/artificial-intelligence>

<sup>14</sup> <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>

<sup>15</sup> <https://www.china-briefing.com/news/artificial-intelligence-china-shenzhen-first-local-ai-regulations-key-areas-coverage/>; but see also <https://merics.org/en/report/lofty-principles-conflicting-incentives-ai-ethics-and-governance-china>.