Editorial: The Central Role of States for Building a Balanced AI Governance

The disruptive nature of artificial intelligence transforms almost all human activities and requires a cohesive and sustainable AI governance framework on a global scale. This framework should aim at managing both the opportunities and the risks derived from this technology in a proportionate manner. The digital economy has increased the need for a trusted ecosystem, including reinforced regulations and additional constraints for all actors dealing with artificial intelligence at whatever part of the value chain. As a result, public actors have initiated a process that promotes a balanced approach beneficial for all innovation, society and individuals. It is part of a concerted international framework at EU, OECD and G20 level and also includes isolated projects like the 'Model AI Governance framework' from Singapore.

The goodwill of States is key in ensuring an effective governance of artificial intelligence. The peer review mechanism or reviews by independent experts can play a central role in the effective implementation of these frameworks. Depending on how AI will be used, it can indeed either contribute to achieving the UN's Sustainable Development Goals (SDGs) or lead to negative societal externalities like harm to citizens, misuse of data, the manipulation of people (deep fake misuse) or mass surveillance.

Within the framework of their sovereignty of positive responsibility and protection, States are responsible for the implementation of these non-binding principles or guidelines on artificial intelligence at a national level. AI governance safeguards users' interest of digital services and products, as well as citizens' interests in public spaces. One of the most recent examples is the EU project to ban facial recognition technologies for up to 3 to 5 years, following the Clearview scandal. This ban is founded on the General Data Protection Regulation and the right 'not to be subject of a decision based solely on automated processing, including profiling, which produces legal effects'.

While the constraint on States to implement the AI principles and guidelines as coordinated at the international level has not changed in nature, this pressure – mainly political – seems to have increased, as is the case in non-cooperative territories in tax matters.

AI Governance Should Consider the Long Term Perspectives

In the wrestling match between excess of individual freedom and the 'common good', the question as to what constitutes 'meaningful governance' is a pertinent one. The AI Transparency Institute holds the opinion that meaningful governance via a binding, directly enforceable regulation is necessary to ensure the safety of AI. It should be proportionate, based on a risk-based approach and respect democratic values and prin-

ciples. Integrating long term criteria within decision making processes would also contribute to mitigating risk and ensuring irreversible steps are avoided.

In particular, ex-ante and ex-post mechanisms should be built for a sustainable AI governance, mainly based on accountability, transparency, good design, safety and liability. Regular risk assessments for high risk projects prior to market deployment of digital products and services and legally binding instruments should be put in place to safeguard the democratic use of artificial intelligence. Soft law like quality labels and certification mechanisms could complete this framework.

In a global market economy, not bound by territorial borders and mainly driven by interdependencies and short term indicators like stock exchange value or polling data, long term interest of future generations should be a key component of the AI governance framework.

Call for Multilateral Governance, Including all Private and Public Stakeholders

Our focus in this Special Issue is on the governance of AI and on Governance via AI. We recommend a hybrid governance methodology by State (hard law) and by the market (soft law), inspired by the GDPR and the 108+ Convention, in particular a network of independent control authorities and effective legal remedies (class actions).

In the first contribution, *Mael Pegny* highlights the need to recognise a right of explanation. His article offers a plea in favour of the transparency of automated decision-making as a requirement for a sustainable trust in a quantified and data-driven society. As he puts it, algorithms and training data scrutiny and auditability are cornerstones of trustworthy AI.

With *Johan Rochel* and *Jean-Henry Morin* we also present a Digital Responsibility Index to quantify the responsibility of economic actors.

In his contribution, *Lexo Zardiashvili* investigates why and how to develop a responsible use of AI within police services and build a groundwork for hard regulation in the law enforcement environment of the Netherlands.

Peter J. Scott analyses historical failures of artificial intelligence and proposes a classification scheme for categorising future failures, while *James D. Miller* shows how time-inconsistency increases the challenge of building an AGI aligned with humanity's values.

Nicolas Miaillhe proposes an analysis of specific use cases, to achieve Sustainable Development Goals (SDG) and formulates proposals for multi-stakeholder collaboration and new kinds of 'public-private-people' partnerships which will reconcile technical, ethical, legal, commercial, and operational frameworks. He advances new international initiatives, such as the Global Data Access Framework and the AI4SDG Center spearheaded as part of a wider international partnership called AI Commons.

Frederic Marty's contribution deals with the governance of platforms. He shows that an increasing use of AI can substantially improve performance in several areas and improve the level of trust in platforms and advanced user dissatisfaction detection tools.

Finally, *Nadisha-Marie Aliman* addresses the complexity of AI governance with safety-relevant, ethical and legal implications at an international level. She also provides novel constructive recommendations for an SDG informed AI governance and an AI-

assisted approach to the SDG endeavour. AI governance could aim at a sustainable transdisciplinary scientific approach instantiated within a corrective socio-technological feedback-loop. She emphasises the need of a strong education and appropriate institutions for the realisation of this potentially robust AI governance strategy.

A Continuously Improving AI Governance Legal Framework

This Special Issue is a first contribution to the discussion of a meaningful AI governance legal framework. It proposes an overview of some of the multifaceted aspects of this topic as well as some concrete proposals to policy-makers as a way forward towards an effective human-centred AI governance. The framework that will be shaped, mainly based on bilateral and multilateral agreements between States, will require continuous improvement.

Eva Thelisson
Guest Editor
AI Transparency Institute