Artificial intelligence: international convention based on "human approach"

The Council of Europe's 'Framework Convention on Artificial Intelligence and human rights, democracy and the rule of law' is about to come into force. This landmark document is the first legally binding international treaty aimed at ensuring respect for human rights, democracy and the rule of law in the use of artificial intelligence systems. The Convention is the result of two years of work by the Committee on Artificial Intelligence (CAI), an intergovernmental body that brought together delegates from 46 Council of Europe member states and 11 observer states (Argentina, Australia, Canada, Costa Rica, Japan, Israel, Mexico, Peru, the Holy See, the United States of America and Uruguay), along with the European Union and more than 70 observers from the private sector, civil society and academia. It was adopted last May in Strasbourg during the annual ministerial meeting of the Council of Europe's Committee of Ministers (i.e. the foreign ministers of the organisation's 46 member states) at the CoE Justice Ministerial Meeting in Vilnius. The Convention has been formally opened for signing today, Thursday 5 September (it needs at least five ratifications to take effect, including at least three by Council of Europe member countries). Non-European countries can also join the Convention. It aims to ensure that activities within the lifecycle of artificial intelligence systems are fully consistent with human rights, democracy and the rule of law, while being conducive to technological progress and innovation. The Framework Convention covers the use of AI systems by public players – including companies acting on their behalf - and private actors. It requires signatories to take measures to identify, assess, prevent and mitigate the risks posed by the use of AI systems, and to determine which measures should be taken when the systems pose risks that are incompatible with international human rights, democracy and rule of law standards. The text, which is currently only available in English and French on the Council's website (https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=0900001680afb11f), consists of 36 articles divided into 8 main chapters (General Provisions, General Obligations, Principles related to activities within the lifecycle of AI Systems, Remedies, Assessment and Mitigation of Risks and Adverse Impacts, Implementation of the Convention, Follow-up Mechanism and Cooperation, and Final Clauses). Most importantly, the "human-centred" approach is the cornerstone of the Convention. Indeed, Article 7 of the Convention states that Parties "shall adopt or maintain measures to respect human dignity and individual autonomy with respect to activities within the lifecycle of artificial intelligence systems." In addition, transparency and oversight - whereby the operations and mechanisms of AI systems must be understandable and accessible - accountability, equality and non-discrimination, the right to privacy and personal data protection, as well as reliability and safe innovation are core principles of the convention. Moreover, the Convention obliges signatory states to establish effective guarantees and safeguards for victims of human rights violations in connection to the use of AI systems. With respect to safeguarding the integrity of democratic processes, the Convention requires States Parties to take measures to ensure that AI systems are not used to undermine the integrity, independence and effectiveness of democratic institutions and processes, including the principle of separation of powers, respect for the independence of the judiciary and access to justice. Activities of adhering States related to the protection of national security interests, national defence, research and development of AI systems are excluded from the scope of the agreement, unless they interfere with human rights, democracy and the rule of law. A Conference of the Parties is tasked with overseeing the implementation of the Convention, but each Party is encouraged to establish an independent oversight mechanism to monitor compliance with the Convention, raise public awareness, and promote public debates and discussions on the use of AI technology. "To negotiate a binding instrument in such an inclusive process and within a very short deadline has been intense and sometimes challenging", said **Thomas Schneider**, (Switzerland), chair of the CAI. "We had to find ways to bridge differences between States' legal systems and

traditions including some difference in the interpretation of some human rights. We also had to manage a range of expectations about how to develop a legal instrument which would have an impact on AI governance at global level." The next step is the development of a methodology for the risk and impact assessment of AI systems on human rights, democracy and rule of law. Critics of the Convention have pointed out that the text was watered down in two aspects during the negotiation process: First of all, the entire issue of national security has been removed from the scope of the Convention. The second critical aspect is the double standard for public and private actors, whereby the former are legally bound, while with respect to the private sector it is left to the Parties to decide how to apply the principles and constraints to the activities of private actors. It has been argued that these compromises were necessary in order to meet the objective of establishing the first international regulation on artificial intelligence that would also apply to countries outside the EU. According to the Secretary General of the Council of Europe, **Marija Pej?inovi?**, the Convention is "a response to the need for a regulation under international law, endorsed by states from different continents that share the same values, which will make it possible to reap the benefits of artificial intelligence while mitigating the risks it poses." The EU's adoption of its own AI law last December confirms that the Convention responds to a specific demand. The UN is also taking action: the 'Summit for the Future', convened by Secretary-General Antonio Guterres in New York on 22-23 September, will bring together world leaders to refocus their efforts on achieving the Millennium Development Goals. At the Summit, world leaders will adopt the Pact for the Future, which will include a Global Digital Compact and a Declaration on Future Generations (Father Paolo Benanti, a Franciscan priest and Professor of Bioethics and Ethics of Technology at the Pontifical Gregorian University in Rome, is a member of the Advisory Board appointed by the UN to formulate recommendations for the Global Digital Compact and on the subject of Artificial Intelligence). The final text of the Summit is not yet public. There have been drafts, and there have been difficulties in negotiating between drastically conflicting approaches, such as those of China and the US. Africa is in the process of developing its own regulatory framework that "reflects its unique context to ensure that the development and use of AI is responsible, ethical and beneficial to all," ConsumerConnect, an independent Nigerian publication, reported in an editorial at the end of August. Rules are more than necessary, harmonising approaches is challenging. The logic of the lowest common denominator is at issue, which is very risky in a field as sensitive as artificial intelligence and its applications.

Sarah Numico