IFRI
STUDIES

GEOPOLITICS OF TECHNOLOGY CENTER

FEBRUARY
2025

Artificial Promises or Real Regulation?

Inventing Global AI Governance

Laure de ROUCY-ROCHEGONDE

The French Institute of International Relations (Ifri) is a research center and a forum for debate on major international political and economic issues. Headed by Thierry de Montbrial since its founding in 1979, Ifri is a non-governmental, non-profit foundation according to the decree of November 16, 2022. As an independent think tank, Ifri sets its own research agenda, publishing its findings regularly for a global audience.

Taking an interdisciplinary approach, Ifri brings together political and economic decision-makers, researchers and internationally renowned experts to animate its debate and research activities.

ISBN: 979-10-373-0982-2

Cover: Image produced on Canva by Emma Badaoui © Ifri, 2025

How to cite this publication:
Laure de Roucy-Rochegonde, "Artificial Promises or Real Regulation? Inventing Global AI Governance", *Ifri Studies*, Ifri, February 2025.

Ifri

27 rue de la Procession 75740 Paris Cedex 15 – FRANCE

Tel.: +33 (0)1 40 61 60 00 – Fax: +33 (0)1 40 61 60 60

E-mail: accueil@ifri.org

Website: Ifri.org

Author

Laure de Roucy-Rochegonde has been the director of Ifri's Center for Geopolitics of Technology since February 2024. Prior to this, she conducted research for the Security Studies Center since 2019, where she investigated the military applications of artificial intelligence, normative conflict and arms control. She holds a doctorate in political science and is an associate researcher at the Center for International Studies (CERI, Sciences Po/CNRS). Her first book, entitled *La Guerre à l'ère de l'Intelligence artificielle : quand les machines prennent les armes* (PUF), was published in October 2024.

Alongside her research activities, she teaches courses on the ethics of war and arms control at Sciences Po Paris and the University of Paris 2 Panthéon-Assas. She also holds a master's degree in public policy and a bachelor's degree from Sciences Po Paris, during which she spent a year in the War Studies department at King's College London.

Executive summary

Arising from the political and economic ambitions of a plurality of players with often divergent interests, the international framework of artificial intelligence (AI) is an acute reflection of contemporary geopolitical tensions.

The risks inherent to the development and mass adoption of AI, a key technology and vector of profound transformations within societies for health, education, employment, or the environment, underline the pressing need to harmonize governance efforts at the international level.

Global governance of AI relies on the ability of state and non-state players to set common standards on technological risks, the boundaries to be drawn, and the principles to be safeguarded. These endeavors aim to promote the safe development of AI that is universal, adapted to cultural diversities, free from bias, and respectful of democratic values and fundamental rights and freedoms.

However, residual political, economic, and legal challenges exacerbated by the limits of existing regulatory frameworks, in the face of increasing balkanization of governance approaches and fragmentation of the international community, considerably complicate the implementation of such an initiative.

Given the intrinsically evolving nature of AI, it is vital to build an adaptable and flexible "future-proof" governance framework capable of anticipating and adjusting to technical advances.

The Summit for Action on Artificial Intelligence to be held in Paris in February is an unprecedented, timely occasion to agree on a shared vision of AI governance that is sustainable and inclusive. For decision-makers, it's an opportunity to better grasp the evolution of practices, regulatory shortcomings, the interests influencing the agreements under construction, and the compromises needed to frame AI on a global scale in the years to come.

Résumé

Fruit des ambitions politiques et économiques d'une pluralité d'acteurs aux intérêts souvent divergents, l'encadrement international de l'Intelligence artificielle (IA) reflète avec acuité les tensions géopolitiques contemporaines.

Les risques inhérents au développement et à l'adoption massive de l'IA, technologie clé et vecteur de transformations profondes au sein des sociétés, pour la santé, l'éducation, l'emploi ou l'environnement, soulignent l'urgence d'harmoniser les efforts de gouvernance à l'échelle internationale.

La gouvernance mondiale de l'IA repose sur la capacité des acteurs étatiques et non étatiques à établir des normes communes sur les risques technologiques, les limites à établir, ainsi que les principes à garantir. Ces efforts visent à promouvoir un développement sécurisé de l'IA, universel, adapté aux diversités culturelles, exempt de biais, et respectueux des valeurs démocratiques ainsi que des droits et libertés fondamentaux.

Cependant, des défis politiques, économiques et juridiques résiduels exacerbés par les limites des cadres réglementaires existants – face à une balkanisation croissante des approches de gouvernance et à la fragmentation de la communauté internationale – complexifient considérablement la mise en œuvre d'une telle initiative.

Par ailleurs, compte tenu de la nature intrinsèquement évolutive de l'IA, il est indispensable que soit élaboré un cadre de gouvernance adaptable et flexible, « future-proof », à même d'anticiper les avancées techniques et de s'y ajuster.

La tenue du Sommet pour l'action sur l'Intelligence artificielle à Paris, début février, est une opportunité sans précédent de s'accorder sur une vision partagée de la gouvernance de l'IA, durable et inclusive. Pour les décideurs, c'est l'opportunité de mieux saisir l'évolution des pratiques, des insuffisances réglementaires, des intérêts qui influencent les accords en construction, et des compromis nécessaires pour encadrer l'IA à l'échelle mondiale dans les années à venir.

Table of contents

Introduction

On February 10 and 11, 2025, the Artificial Intelligence Action Summit will be held in Paris, bringing together heads of state and government, representatives of international organizations, business leaders, academics, non-governmental organizations (NGOs), artists, and members of civil society from over 100 countries. This summit, jointly chaired with India, continues the work of the Artificial Intelligence Safety Summit, held at Bletchley Park (United Kingdom) in November 2023 under the leadership of the UK government, with two subsequent events in Seoul in May 2024 and San Francisco in November 2024. The shift in focus for the French summit, now dedicated to action rather than security, highlights the tension between the risks and opportunities presented by artificial intelligence (AI), which makes international regulation of the development and use of this technology so difficult.

The term "artificial intelligence" was coined in 1956 by logician John McCarthy to describe methods to better understand and imitate human intelligence through computer programs.[1] As defined by the European Union's (EU) High-Level Expert Group on AI, set up by the European Commission in June 2018, the term refers to "systems that display intelligent behavior by analyzing their environment and taking actions – with some degree of autonomy – to achieve specific goals".

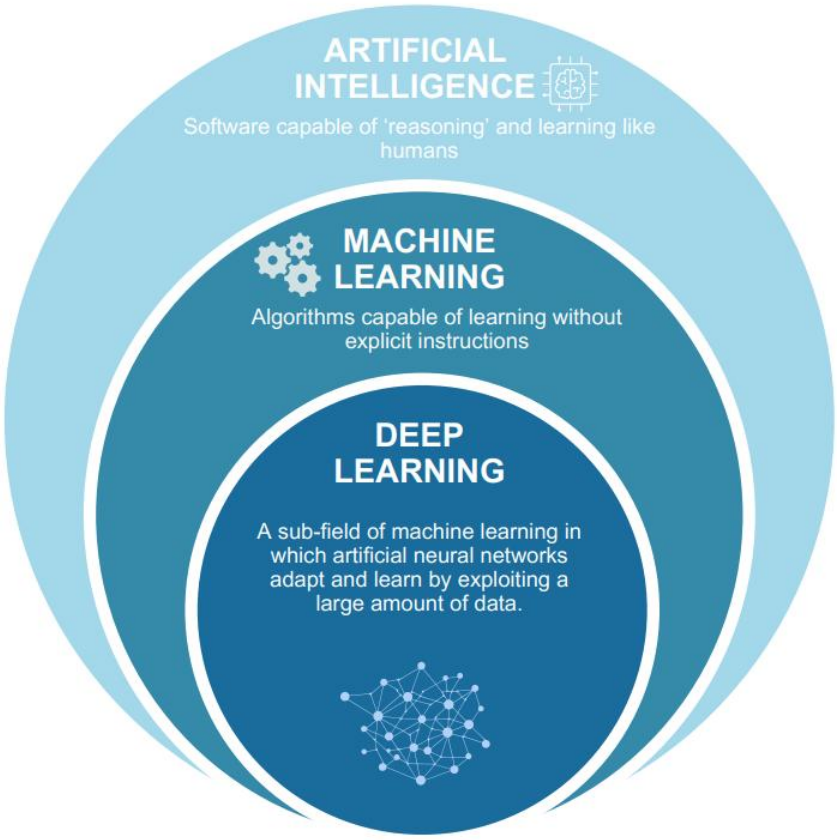
AI is often divided into two basic technical categories:

knowledge representation and reasoning (symbolic AI), which structures knowledge formally for algorithms to use in producing their output;

data-driven methods, in particular machine learning (connectionist AI), which are also used in generative AI models.[2]

1. J.-C. Noël, "Intelligence artificielle : vers une nouvelle révolution militaire ?", *Focus stratégique*, No. 84, Ifri, October 2018.
2. The latter is the result of advances in deep learning and natural language processing, which enable computer systems to "understand" human language and perform complex tasks, including the automated generation of text or visual content from prompts.

Schematic representation of the different artificial intelligence techniques



AI systems are ever more widely present in our daily lives and, more generally, in the global economy. The commercial success of ChatGPT, which had attracted 100 million users only a few weeks after its launch at the end of 2022, is the clearest illustration of this new trend. Two years later, OpenAI's chatbot now has over 300 million active weekly users.[3]

Since 2023, AI has received significant media coverage. While its potential benefits are endlessly highlighted – better healthcare, safer and more environmentally friendly transportation, increased productivity, cheaper and more sustainable energy – its risks are also regularly addressed. In May 2023, for example, the "godfather of AI", Geoffrey Hinton – who last year received the Nobel Prize in Physics for his "foundational discoveries and inventions that enable machine learning with artificial neural networks" – announced his resignation from Google, sounding the alarm on the dangers inherent in his field of research. In an interview with the New York Times, he warned that "future versions of the

3. E. Roth, "ChatGPT Now Has over 300 Million Weekly Users", *The Verge*, December 4, 2024.

technology pose a threat to humanity". He expressed concern about the growing sophistication of AI systems and the risk of them falling into the wrong hands. "A part of [me] regrets [my] life's work",[4] he also confessed.

AI is in the midst of an all-out power struggle, unsurprisingly dominated by Sino-American rivalry, though other nations, such as France, the United Kingdom, Germany, India, Canada, South Korea and the Gulf States, have also been vying for a stake in the game.[5] In such a context of intense political, geopolitical, and economic competition, perspectives and discourse diverge regarding what should take priority in terms of AI governance.[6] While some stress the need for regulation, others insist that innovation should be fostered and regard attempts at regulation as an obstacle to progress. These contrasting perspectives reflect profoundly divergent national priorities, tied to differing interests and attitudes towards norms and regulations.

Given that AI technologies are part of transnational value chains, and as they are becoming increasingly accessible and exponentially widespread, their regulation can only be considered at the international level. However, in addition to the usual stumbling blocks of multilateralism, the technical specificities of AI make its oversight all the more complex. How, then, should the international governance of AI be approached?

Global governance, which refers to the "collective effort by sovereign states, international organizations and other nonstate actors to address common challenges and seize opportunities that transcend national frontiers,"[7] is a familiar topic of debate in international relations. Central to this controversy is the question of implementing universal regulations for global security issues.

Since 2019 – and with increasing urgency since 2023 – initiatives to regulate AI have multiplied, promoted by a wide range of actors, from governments to international and regional organizations, as well as business coalitions and citizens' associations. In late October 2023, the G7 adopted a non-binding code of conduct for AI developers.[8] A few days later, in November, the United Kingdom hosted the Bletchley Park Summit; meanwhile, after signing an executive order aimed at promoting "safe, secure, and trustworthy" AI, U.S. President Joe Biden met with his Chinese counterpart Xi Jinping to initiate a bilateral dialogue on the military uses of AI. European legislators, for their part, adopted the AI Act at the start of

4. C. Metz, "The Godfather of AI Quits Google and Warns of Danger Ahead", *The New York Times*, May 4, 2023.
5. B. Pajot, "Artificial Intelligence, or The Race for Power", *Politique étrangère*, Vol. 89, No. 3, Ifri, September 2024.
6. J. B. Bullock et al. (eds.), *The Oxford Handbook of AI Governance*, Oxford: Oxford University Press, 2022.
7. St. Patrick, "The Unruled World: The Case for Good Enough Global Governance", *Foreign Affairs*, Vol. 93, No. 1, Winter 2014.
8. R. Balenieri, "IA: les pays du G7 adoptent un code de bonne conduite", *Les Échos*, October 30, 2023.

2024. This landmark text, which came into force in August of the same year, aims to regulate the risks associated with this technology while serving as a global benchmark for regulation.

However, this fragmented landscape runs the risk of engendering inconsistent governance frameworks, uncoordinated, overlapping dialogues, and divergent collective priorities, which could compromise innovation and hinder the development of AI for the common good. The objective of this paper is therefore to understand the obstacles to IA governance, so that they can be overcome. To begin, we will demonstrate the importance of a global approach to AI oversight. We will then examine the current "balkanization" of its governance. Finally, we will explore avenues for better international regulation of AI.

A global governance concern

Recent advances in Large Language Models (LLMs)[9] and chatbots such as ChatGPT (OpenAI), Gemini (Google), and Ernie Bot (Baidu) have democratized the use of generative AI. Using these tools, the general public familiarized itself with AI's many applications, which simultaneously generated a great deal of enthusiasm and concern.

Beginning in 2023, researchers and public officials became increasingly vocal about the dangers of this technology, warning that it could lead to layoffs, threaten democracy, infringe on civil liberties and privacy, and jeopardize intellectual property and copyright protections.[10] The necessity and urgency to establish regulations to ensure AIs' lawful development and use have become a recurring theme in the current public discourse. In addition to national security and human rights concerns, however, there is also the need to maintain economic competitiveness. Many actors in this field are strongly opposed to the idea of "tying our hands" with regulations, while other powers could make unrestricted use of this technology.

AI, due to its enormous economic, political, and social impact, raises many challenges when it comes to governance. While these issues were initially addressed by national authorities, initiatives led by international organizations have gained considerable ground in recent years.[11] This first part, therefore, aims to demonstrate the need for a regulatory framework for AI development and deployment and why such efforts can only be effective on a global scale.

Mitigating the risks of AI

Why has AI become a priority for regulators throughout the world? The answer may lie in how the risks associated with this technology have been highlighted.

From Vladimir Putin's declaration, as early as 2017, that "whoever becomes the leader in [AI] will become the ruler of the world", to Elon Musk's assertion that "AI is far more dangerous than nukes" and will be the "most likely cause of WW3", warnings about the inherent dangers of AI

9. LLMs are foundation models trained on vast amounts of data to understand and generate natural language text and other types of content, in order to accomplish a wide range of tasks.
10. M. Schaake, "The Premature Quest for International AI Cooperation", *Foreign Affairs*, December 21, 2023.
11. Ibid.

have indeed multiplied in recent years. However, as philosopher and historian Émile Torres notes, these oft-repeated dystopian visions distract from very real issues:

> "Talking about human extinction, about a genuine apocalyptic event, [...] is just so much more sensational and captivating than Kenyan workers getting paid US$1.32 an hour [to moderate content used by AI], or artists and writers being exploited [to train these systems]."[12]

Discourse on the risks posed by AI is indeed something of a competitive sport.[13] The digital giants urge us to focus on sometimes outlandish long-term threats (even fearing the extinction of the human species)[14], obscuring more immediate and tangible dangers (concerning intellectual property or taxation, for example). Nevertheless, the calls for oversight directly result from such threats being considered. What, then, are the main risks associated with AI that would require regulation?

For the "common good"

Firstly, the spectacular development of AI raises many concerns over the protection of the "common good", regarding ethical, social, economic, and environmental issues.[15]

As noted by the United Nations (UN) Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance, Ashwini K.P., "[r]ecent developments in the field of generative artificial intelligence and the burgeoning application of artificial intelligence continue to raise serious human rights issues, including concerns about racial discrimination".[16] AI systems tend to reproduce and amplify the biases present in training corpora because the datasets used to train the algorithms are generally incomplete, with certain groups of people being underrepresented.[17] Such over- or under-representation of particular groups in learning datasets, namely when it is based on ethnic criteria, generates algorithmic biases. Similarly, if the input data is already biased –

12. In his work, Émile Torres highlights how notions of "AI safety" and the "pursuit of human well-being" are used by certain schools of thought to influence AI development priorities and avoid accountability to the wider public. See É. Torres, *Human Extinction: A History of the Science and Ethics of Annihilation*, London: Routledge, 2023.
13. B. Pajot, "Les risques de l'IA : enjeux discursifs d'une technologie stratégique", *Études de l'Ifri*, Ifri, June 2024.
14. K. Roose, "A.I. Poses 'Risk of Extinction', Industry Leaders Warn", *The New York Times*, May 30, 2023.
15. M. Coeckelbergh, "Artificial Intelligence, the Common Good, and the Democratic Deficit in AI Governance", AI and Ethics, 2024.
16. A/HRC/56/68, "Report of the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance", Human Rights Council, United Nations General Assembly, June 3, 2024.
17. E. Ferrara, "The Butterfly Effect in Artificial Intelligence Systems: Implications for AI Bias and Fairness", *Machine Learning With Applications*, Vol. 15, 2024.

because it comes from the web, for example, where racist or sexist discourse is common – the algorithms will logically produce biased results.[18]

These biases can have serious consequences for marginalized individuals and communities, fueling discrimination.[19] The most salient examples of this phenomenon are automated predictions concerning prisoners' potential for recidivism, which have had an impact on parole decisions; certain algorithms used to assist in recruitment have systematically favored men over women.[20] A study of image databases used by law enforcement agencies in the United States has revealed that African-Americans are more likely to be wrongly incriminated by facial recognition systems.[21] This bias reflects the overrepresentation of African-Americans in police image databases.

Another common form of bias in AI tools results from how they are developed. Even if the data used to train an algorithm is perfectly representative, decisions made during its development can result in distorted results and generate significant discriminatory effects. For example, when developing an algorithm to assess credit risk, the definition and measurement of vulnerability factors can lead to biased results.[22] Using credit scores as the main indicator can, therefore, disadvantage those groups of people whose credit scores tend to be lower.

More generally, the automation of certain forms of arbitration – involving employment, healthcare, bank loans, or education programs – raises ethical questions regarding the transparency and accountability of these systems. One of AI's major goals is, in fact, decision-making free of human intervention, which is often seen as occurring in a "black box". Certain AI-based programs can, for example, make decisions autonomously, continuously updating themselves as they are exposed to new data. Yet these updates can cause them to rely on criteria that gradually diverge from those that were initially programmed, based on the trends identified in the data.

As these new trends influence algorithms' decisions, it becomes increasingly difficult for users to discern what factors contribute to the results they receive. This lack of transparency only serves to make the systems' reasoning processes even more elusive and opaque.[23] Many programs developed by private firms are also exempt from all external legal scrutiny,

18. Shortly after launching its first chatbot, Tay, in 2016, Microsoft was forced to suspend its social media interactions to stem the flow of racist and sexist posts.
19. N. Mehrabi et al., "A Survey on Bias and Fairness in Machine Learning", *ACM Computing Surveys*, Vol. 54, No. 6, 2022.
20. J. Angwin et al., "Machine Bias", ProPublica, May 13, 2016.
21. Cited in C. O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, London: Penguin Books, 2017.
22. A/HRC/56/68, "Report of the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance", op. cit.
23. J. Burrell, "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms", *Big Data & Society*, Vol. 3, No. 1, 2016.

due to contract and intellectual property laws. This lack of oversight exacerbates concerns regarding accountability and further complicates efforts to regulate these technologies.

Privacy is another key issue.[24] AI systems often rely on data containing personal information, and the collection or processing of such data without users' consent is a privacy violation.[25] Furthermore, data that was initially collected for a specific purpose, such as healthcare (in the case of medical applications, for example), can find itself circulating without the consent of the individuals concerned and used outside of its original scope, such as in the legal system. Additionally, data leaks and unauthorized access, including through computer hacking, raise new concerns around the protection of personal information.[26]

Such invasions of privacy are all the more concerning given that AI tools can also serve as instruments of social control and mass surveillance.[27] According to the AI Global Surveillance Index, in 2019, 56 countries had already introduced AI into their urban surveillance systems. Autocratic states can use AI-based surveillance to identify and track individuals to preemptively suppress civil disobedience, thereby further entrenching their authority. In China, for example, the government uses such systems to police ethnic minorities, among them Uyghurs, as part of its Sharp Eyes Program.[28] Meanwhile, U.S. company Clearview AI collected billions of images of individuals through social media without their permission, resulting in lawsuits in the United States and Europe.[29] Such practices pose a serious threat to personal data protection and citizens' fundamental rights.

Thirdly, the combined effects of AI, automation, and robotics in many areas of activity are rapidly transforming today's workplace.[30] A 2024 report by strategy consulting firm McKinsey predicted that up to 45% of hours worked could be automated by 2035, in particular in sectors such as logistics, accounting and healthcare.[31] The same year, Goldman Sachs

24. Generative AI could quickly render the current legal framework for the protection of personal data obsolete given the complexity of its application. Instances of non-compliance with the General Data Protection Regulation (GDPR) have multiplied in Italy, Spain and France, for example.
25. S. Lai and B. Tanner, "Examining the Intersection of Data Privacy and Civil Rights", Brookings Institution, July 18, 2022.
26. J. King and C. Meinhardt, "Rethinking Privacy in the AI Era: Policy Provocations for a Data Centric World", *White Paper*, HAI, Stanford University, February 2024.
27. A. Olvera, "How AI Surveillance Threatens Democracy Everywhere", *Bulletin of the Atomic Scientists*, June 7, 2024.
28. D. Peterson, "How China Harnesses Data Fusion to Make Sense of Surveillance Data", Brookings Institution, September 23, 2021.
29. F. Filloux, "Reconnaissance faciale : Clearview AI, le poison mortel de la vie privée", *L'Express*, October 18, 2023.
30. D. Acemoglu and P. Restrepo, "The Wrong Kind of AI? Artificial Intelligence and the Future of Labour Demand", *Cambridge Journal of Regions, Economy and Society*, Vol. 13, No. 1, 2020; J. Nocetti, "Europe Facing the Digitization of Work: The Political Risks", *Études de l'Ifri*, Ifri, September 2018.
31. D. Barroux, "Presque la moitié des heures travaillées peuvent être automatisées par l'IA", *Les Échos*, May 24, 2024.

estimated that, worldwide, 300 million full-time jobs were at risk.[32] Such widespread layoffs are likely to increase social inequalities, as low-skilled workers – who perform the most repetitive, and therefore automatable, jobs – will be the first to be affected.[33]

The rapid development of these technologies, therefore, risks exacerbating already significant disparities, with far-reaching sociopolitical repercussions. Such an impact on employment would be all the more damaging as society struggles to implement sustainable solutions to provide for its most vulnerable or marginalized members,[34] who would find themselves falling even further behind. Certain Silicon Valley magnates, such as Sam Altman or, before him, Bill Gates and Mark Zuckerberg, have called for the adoption of a universal basic income, i.e., a minimum pension that would mitigate the effects of job losses and the potential social unrest that would result from the sudden impoverishment of a portion of the population.[35]

Fourth, the issue of intellectual property is a significant concern, as generative AI upends any traditional understanding of copyright.[36] Indeed, LLMs are trained on data that is gathered online, often without consent or compensation for its creators. This then raises the question of whether the large-scale use of data to train LLMs can be considered as an instance of 'fair use', as the major players in generative AI claim, in particular, OpenAI.[37] This principle of Anglo-Saxon copyright law allows for certain use cases without requiring authors' prior consent, provided that the work is substantially transformed. This question, therefore, applies to automatically generated content that reproduces the style of specific artists, which constitutes copyright infringement.

This is why, since 2023, numerous artists, authors, comedians, developers, record companies, and media groups have filed lawsuits against a number of American digital giants, including OpenAI, Microsoft, Stability AI, Midjourney, Meta, and Anthropic, to name but a few. In December 2023, the New York Times filed a lawsuit against OpenAI and Microsoft for using the content of their articles without their consent to train ChatGPT and Copilot.[38] They also accuse them of having reproduced

32. N. Beyler, "ChatGPT et l'IA menacent 300 millions d'emplois dans le monde, selon Goldman Sachs", *Les Échos*, March 28, 2023.
33. White-collar workers whose jobs are also affected can more easily adapt their careers to keep up with technological change.
34. K. Georgieva, "AI Will Transform the Global Economy: Let's Make Sure It Benefits Humanity", IMF Blog, January 14, 2024.
35. OpenAI estimates that more than 80% of workers in the United States could see their jobs affected by AI. See S. Emerson, "OpenAI Founder Sam Altman Gave Thousands of People Free Money. Here's What Happened", *Forbes*, July 22, 2024.
36. C. Metz, "Lawsuit Takes Aim at the Way A.I. Is Built", *The New York Times*, November 23, 2022.
37. "Does Generative Artificial Intelligence Infringe Copyright?", *The Economist*, March 2, 2024.
38. "Le *New York Times* poursuit en justice Microsoft et OpenAI, créateur de ChatGPT, pour violation de droits d'auteur", *Le Monde*, December 27, 2023.

several articles in full – unaltered – giving specific examples. These practices, according to the newspaper, exceed the generally accepted limits of the doctrine of fair use. The newspaper also argues that this damages its relationship with its readers and negatively impacts its main sources of income, including subscriptions, advertising and partnerships. These cases highlight the need for a clear legal framework to protect creators while promoting innovation.

Finally, AI has colossal environmental costs. The data centers used to train and operate LLMs consume enormous amounts of energy. The popularity of these models leads to significant resource consumption, from their development to their use. Manufacturing microchips, storing data, training models, processing user requests, and producing data all have a significant impact on physical, water, and energy resources, with direct repercussions on the climate.

According to a study by the University of Massachusetts, the training required to develop just one of these LLMs can emit as much carbon dioxide (CO_2) as five cars do in their lifetime.[39] The water requirements for cooling data centers have not been made public, but they are likely to be considerable as well.[40] The implementation of a harmonized global standard for transparency, overseen by national governments, would make environmental data accessible to researchers and journalists. Such an approach would allow for public scrutiny of AI companies' consumption of natural resources and enable policymakers to set relevant and effective constraints. AI governance, therefore, cannot be dissociated from environmental protection efforts.

For international peace and stability

A second area of major concern is the threat that AI poses to international peace and stability. In this area, cybersecurity risks are particularly significant. Cyber-attackers have already begun to take advantage of LLMs' capabilities to produce computer code, perform automatic translations and curate technical content. Such practices, which have already been observed in Chinese, Iranian, North Korean, and Russian groups,[41] multiply their capabilities in terms of social engineering, identity theft and manipulation through better targeting and more sophisticated phishing techniques.[42]

39. K. Haro, "Training a Single AI Model Can Emit as Much Carbon as Five Cars in Their Lifetimes", *MIT Technology Review*, June 6, 2019.
40. They were also implicated in the Los Angeles wildfires at the start of 2024, due to the depletion of water resources. See T. Katzenberger, "AI Data Centers Face Scrutiny for Water and Energy Use as LA Fires Rage", Politico, January 9, 2025.
41. "Staying Ahead of Threat Actors in the Age of AI", Microsoft, February 14, 2024, available at: www.microsoft.com.
42. J. Hazell, "Spear Phishing with Large Language Models", Oxford Internet Institute, December 14, 2023, available at: www.governance.ai.

In addition, since AI technologies rely on computer systems, they inevitably contain electromagnetic access points through which data can be transmitted to carry out malicious attacks. Their widespread use therefore increases the potential attack surface, both for these systems and the networks in which they operate.[43]

According to the UK's National Cyber Security Centre (NCSC), LLMs are particularly vulnerable to two types of attack:

prompt injections, malicious instructions whose objective is to manipulate the model through its command interface;

data poisoning, datasets that were tampered with either before or during model training.

Because they are compiled from massive open datasets and tend to be used to provide information for third-party applications and services, attacks aimed at corrupting these datasets represent a significant risk.[44]

These new capabilities fuel fears of a proliferation of ransomware and distributed denial of service (DDoS) cyberattacks, as AI technologies also enable better coordinated attacks through networks of compromised systems (botnets). The organized crime industry has already fully embraced these emerging tools,[45] and there are now countless LLMs on the dark web.[46] It is therefore reasonable to assume that state powers that are heavily invested in offensive cyber warfare will take full advantage of this opportunity.

Secondly, and related to the cybersecurity risk, there are growing concerns over the integrity of information in the age of LLMs. Beyond the "hallucinations" observed in chatbots, which are sometimes prone to inventing or distorting information,[47] generative AI raises many fears concerning disinformation. It makes it considerably easier to manipulate information, both in terms of production and dissemination, namely as part of foreign digital interference campaigns. Deepfakes, i.e., fabricated informational content created using generative AI, have proliferated on social networks, with tangible real-world repercussions.

On October 19, 2023, the front page of French newspaper *Libération* featured a photograph of a protester in Cairo brandishing an image of a baby crying in the rubble of the Al-Ahli Arabi hospital in Gaza. In reality, this image was AI-generated and first began circulating following the earthquakes that occurred near the Turkish-Syrian border in February

43. J. Jun, "How Will AI Change Cyber Operations", War on the Rocks, April 30, 2024.
44. B. Pajot, "Les risques de l'IA : enjeux discursifs d'une technologie stratégique", op. cit.
45. D. Larousserie, "Comment les chatbots ont été gangrenés par l'industrie du crime organisé", *Le Monde*, February 13, 2024.
46. J. Cheminat, "Après WormGPT, les cybercriminels livrent FraudGPT", *Le Monde Informatique*, July 26, 2023.
47. C. Metz, "Chatbots May 'Hallucinate' More Often Than Many Realize", *The New York Times*, November 6, 2023.

2023.[48] Though there has been no shortage of photographs of children amid the rubble in Gaza, this fabricated image only served to fuel suspicions of fake news reports and staged Palestinian victims – sometimes referred to as "Pallywood".[49]

Generative AI also facilitates the emergence of new, "complex" cyber-influence tactics. In October 2023, pro-Iranian hackers interrupted a news broadcast in the United Arab Emirates, replacing it with a fake, AI-generated report on the war in Gaza.[50] In Romania, the presidential election of December 2024, which saw the ultra-nationalist and conspiracy theorist candidate Călin Georgescu emerge as the frontrunner after the first round, was annulled due to suspicions that Russia had used TikTok to manipulate the vote in his favor, as the candidate supported the immediate suspension of aid to Ukraine. How this information interference campaign was carried out remains unclear at this stage, but it is likely that generative AI was used to develop and amplify pro-Georgescu content on the Chinese platform.[51] AI thus plays a role in magnifying international tensions, serving the strategic interests of states seeking to infiltrate the information spheres of their adversaries in order to exert influence and manipulate public opinion.

AI tools are therefore making information attacks more massive, more sophisticated, and more targeted, enabling them to reach a much larger target audience while circumventing the detection mechanisms deployed by platforms, through indirect channels, as opinion leaders and influencers inadvertently share inauthentic content. Although they have in the past publicized their determination to combat such attempts at disinformation,[52] Big Tech is regularly criticized for its mismanagement of fabricated content.[53] Such repeated examples of information manipulation demonstrate the need for AI-generated content to be labeled as such, something which, although now required by the European Digital Services Act,[54] is only just being introduced.[55]

48. "'Libé' s'est-il rendu coupable d'une 'fake new' en publiant la vraie photo d'un homme brandissant une image générée par IA ?", *Libération*, October 19, 2023.
49. W. Audureau, "'Pallywood': en plein carnage à Gaza, le mythe des fausses morts palestiniennes", *Le Monde*, December 16, 2023.
50. D. Milmo, "Iran-backed Hackers Interrupt UAE TV Streaming Services with Deepfake News", *The Guardian*, February 8, 2024.
51. M. Bran, "En raison de l'influence de TikTok, les juges roumains annulent la présidentielle", *Le Monde*, December 7, 2024.
52. In February 2024, for example, Google, Meta, Microsoft, OpenAI, TikTok and Adobe signed an agreement to combat deepfakes during elections. See G. De Vynck, "AI Companies Agree to Limit Election 'Deepfakes' But Fall Short of Ban", *The Washington Post*, February 13, 2024.
53. C. Zakrzewski, "ChatGPT Breaks Its Own Rules on Political Messages", *The Washington Post*, August 28, 2024.
54. Pursuant to Article 35-k, and if systemic risks are identified on the platform, such as opinion manipulation, and negative effects on democratic and electoral processes, as well as civic discourse.
55. K. Lentschner, "Les plateformes tâtonnent face à la labellisation des contenus IA", *Le Figaro*, July 2, 2024.

Thirdly, the promise of AI is also the subject of growing expectations and rapid innovation on the battlefield, which raises many questions with regard to strategy, policy, law and ethics.

"The First AI War": this is how Time magazine presented the Russo-Ukrainian war on the cover of its February 26, 2024 issue. The conflicts in Ukraine and Gaza have indeed proven AI to be a true force multiplier: it has a wide range of military applications, from logistics to targeting, as well as intelligence and decision support for command and control (C2) activities.[56] It has even been presented as a new revolution in warfare, as significant as gunpowder or nuclear weapons before it.

Advances in military AI, combined with advances in robotics, do however raise fears of their misuse, in particular concerning lethal autonomous weapons systems (LAWS), which the media have dubbed "killer robots". These are systems that, once activated, can identify a target and use lethal force without human supervision.

This raises the issue of metacognition, were systems to continue to learn while in operation, in order to adapt to changing environments. Without effective supervision, what the system "learns" during operation may result in unexpected and inappropriate behavior outside of its intended operating conditions.[57] More broadly, self-learning systems capable of modifying themselves while in operation, beyond the issue of their configuration, call into question whether it is even possible to guarantee their long-term reliability.

The French strategy for artificial intelligence in support of defense also argues that revisionist powers like Russia and China are investing in military AI to disrupt the international status quo to their advantage. Innovations in this field could indeed level the strategic playing field, as they are relatively inexpensive and easily adopted. Other players could also join the fray by acquiring these technologies which, while complex, are increasingly affordable and therefore ever more accessible. Such widespread access would allow weaker actors to shift the international balance of power.

56. According to Ukrainian weekly *Dzerkalo Tyjnia*, the Ukrainian army currently employs AI in ten different areas: weapons system autonomy; observation and reconnaissance; target identification and classification; threat analysis and prediction; logistics and supply; cybersecurity; electronic warfare; simulation and training; military health; and decision support. On this subject, see A. Férey and L. de Roucy-Rochegonde, "From Ukraine to Gaza: Military Uses of Artificial Intelligence", *Politique étrangère*, Vol. 89, No. 3, Ifri, September 2024.

57. For example, during a simulation exercise conducted by the U.S. Air Force, a drone operated by an AI program allegedly decided to "kill" its operator to stop him from interfering as it attempted to fulfill its mission. A spokesperson for the U.S. Air Force however denied the existence of this simulation. See "US Air Force Denies Running Simulation in Which AI Drone 'Killed' Operator", *The Guardian*, June 2, 2023.

The weaponization of AI therefore generates a great deal of concern. In November 2018, at the first Paris Peace Forum, the Secretary-General of the United Nations (UN), Antonio Guterres, called for such weapons, which he described as "politically unacceptable [and] morally repugnant", to be "prohibited by international law"[58]. In October 2023, together with the president of the International Committee of the Red Cross, Mirjana Spoljaric Egger, he reiterated his call for their prohibition. Concerns include whether such systems could be compatible with the law of armed conflict and the principle of human dignity, the risk of lowering the threshold for conflict and destructive escalations, and their dissemination to violent non-state actors.[59]

AI also carries the risk of lowering the technological barrier limiting access to weapons of mass destruction. AI models could indeed be misused to create new and potentially more dangerous chemical or biological agents than any known formulas, putting them in the hands of unauthorized actors (terrorist groups, criminals, etc.). Malicious state actors could also have the capacity to develop new weapons, making them more lethal and more difficult to identify, hindering the development and deployment of effective countermeasures.

As a result, the regulation of military AI is now the subject of discussions in multilateral negotiation forums. At the international level, the debate on autonomous weapons was initiated by NGOs united in a Campaign to Stop Killer Robots in 2012. It was then taken up by the Human Rights Council, as well as the Convention on Certain Conventional Weapons, which in 2016 decided to create a group of governmental experts tasked with examining the issue, their mandate being renewed in 2023. More recently, during their last official meeting in November 2024, Presidents Xi and Biden agreed on the need to limit the integration of AI in nuclear weapons systems.[60]

There is therefore a need to establish clear limits on the uses of AI-powered weapons, including cyberweapons. The implementation of international law in the context of cyber operations is already poorly defined;[61] AI further complicates the issue. This technology gives cyber-attackers an edge by enabling them, for example, to use generative AI to quickly analyze large volumes of software and identify vulnerabilities. An international agreement prohibiting certain uses of weaponized AI, such as autonomous weapons, or the dissemination of disinformation during

58. "Secretary-General's address to the Paris Peace Forum", Paris, November 11, 2018.
59. L. de Roucy-Rochegonde, *La Guerre à l'ère de l'Intelligence artificielle : quand les machines prennent les armes*, Paris: PUF, 2024.
60. L. Egan and P. Kine, "Biden's Final Meeting with Xi Jinping Reaps Agreement on AI and Nukes", Politico, November 16, 2024.
61. F. Delerue, *Cyber Operations and International Law*, Cambridge: Cambridge University Press, 2020.

election campaigns in another country, could provide essential safeguards and promote responsible practices.

Assigning externalities

Beyond the issue of the risks posed by AI, there is also the question of how to assign the associated externalities. On the one hand, some have criticized the concentration of power within major technology companies in this new "summer" of AI. On the other hand, its development takes place in a transnational context and causes cross-border externalities, requiring global cooperation to develop regulatory frameworks which transcend borders.

Firstly, the foundation models capable of accomplishing a wide range of tasks, on which the secondary models and AI applications the public is familiar with are based, are only developed by a few tech giants. Indeed, the exorbitant costs involved in developing these models tend to favor early movers, resulting in high market concentration.[62] This is further reinforced by the predatory practices of Big Tech, as evidenced by Google's investment in Deepmind, Microsoft's in OpenAI and Mistral AI, and Amazon's in Anthropic.[63]

These industry giants have the strategic resources – funds, semiconductors, computing power, data, algorithms, cloud and talent – necessary to steer the major trends in AI.[64] Their colossal resources also enable them to attract the most promising minds the world over, gutting the academic research community and public institutions in the process,[65] not to mention their willingness to poach the best talent from smaller players like Stability AI,[66] and their well-documented anti-competitive and monopolistic practices.[67] In the words of Benoît Cœuré, President of the French Competition Authority: "AI is the first technology to be dominated by major players from the outset".[68] It is indeed the first disruptive innovation whose development capabilities are entirely controlled by the most powerful companies. This is why a growing number of actors are calling for a policy

62. J. Vipra and A. Korinek, "Market Concentration Implications of Foundation Models: The Invisible Hand of ChatGPT", Brookings Institution, September 7, 2023.
63. E. Ludlow, M. Day and D. Bass, "Amazon to Invest Up to $4 Billion in AI Startup Anthropic", Bloomberg, September 25, 2023.
64. A. Kak, S. Myers West and M. Wittaker, "Make No Mistake, AI Is Owned by Big Tech", *MIT Technology Review*, December 5, 2023.
65. N. Nix, C. Zakrzewski and G. De Vynck, "Silicon Valley Is Pricing Academics Out of AI Research", *The Washington Post*, March 10, 2024.
66. T. Warren, "Stability AI CEO Resigns to 'Pursue Decentralized AI'", *The Verge*, March 23, 2024.
67. D. Milmo, "UK Watchdog to Examine Microsoft's Partnership with OpenAI", *The Guardian*, December 8, 2023.
68. A. Piquard, "L'IA est la première technologie à être d'emblée dominée par les grands acteurs", *Le Monde*, September 27, 2024.

combining regulation and public investment in AI, in order to counterbalance the growing influence of private actors.[69]

Though it is sometimes presented as an alternative to the hegemony of Big Tech, open source is not without its own risks. Granted, "open" innovations are less exposed to commercial pressure, which pushes proprietary developers to work as quickly as possible in order to maintain their competitive advantage, at the risk of their models being neither fully developed nor sufficiently secure. However, precisely because they are open, their extremely powerful tools can be used by anyone, including malicious actors, which is a cause for concern.[70] In November 2024, it was revealed that China had used Meta's Llama 13B model to develop a military chatbot[71]. Open models are also more exposed to cyber risks, particularly in terms of data contamination and poisoning.

There is also a North/South divide in the global distribution of the costs and benefits of AI. Developing and emerging countries are particularly vulnerable to the social upheavals caused by AI, due to the pressure it puts on low-skilled and low value-added jobs. These countries also lack sufficient social security coverage and effective resources to enable them to make a technological leap and truly benefit from the development of AI.[72]

These countries also have a large number of the "click workers" who are essential to the training of AI models.[73] But their extremely impoverished standard of living contrasts sharply with the colossal profits generated by this industry.[74] The asymmetry in data collection, combined with an inequitable distribution of costs and benefits – AI relying on mining resources and labor from the Global South, but designed primarily by and for people in the Global North – contributes to an acute sense of injustice. This imbalance is exacerbated by the blatant lack of representativeness in AI models, which are often trained on data from the Western world, mostly produced by high-income men, and written in English. These systemic biases perpetuate stereotypes that can not only accentuate social inequalities at the national level, but also exacerbate political and cultural differences internationally.[75]

69. M. Schaake, "AI Is Too Important to Be Monopolised", *Financial Times*, December 12, 2024.
70. This concern is often raised by American political and economic actors, as it has the dual advantage of drawing attention to Chinese threats while strengthening the position of Big Tech companies and their closed models.
71. J. Pomfret and J. Pang, "Chinese Researchers Develop AI Model for Military Use on Back of Meta's Llama", Reuters, November 1, 2024.
72. D. Björkegren, "Artificial Intelligence for the Poor: How to Harness the Power of AI in the Developing World", *Foreign Affairs*, August 9, 2023.
73. G. Kristanadjaja, "Intelligence artificielle : dans les pays du Sud, des petites mains victimes de 'colonisation numérique'", *Libération*, March 21, 2024.
74. B. Perrigo, "OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic", *Time*, January 18, 2023.
75. V. Türk, "How AI Reduces the World to Stereotypes", *Rest of World*, October 10, 2023.

The World Health Organization (WHO) has warned of the potentially harmful effects of the use of AI technologies in healthcare for people in developing countries. In particular, it has highlighted the issue of the lack of diversity in training data, which limits its effectiveness for under-represented groups. The organization has also expressed concern about the private sector's dominance, to the detriment of academic research and initiatives by public agencies.

Given this context, calls for the introduction of regulatory mechanisms to avoid exacerbating these issues are becoming increasingly urgent, including at the international level.[76] One of the approaches under consideration is the introduction of a dedicated tax to offset the social impacts of AI.[77] Because of the stark asymmetry in the distribution of the positive and negative externalities of AI,[78] some advocate for "techno-prudential" governance or for the containment of these technologies.[79] The aim is to mobilize decision-makers, industrial stakeholders and actors from civil society to ensure collective control of these tools while ensuring the fair distribution of their externalities on a global scale.

The emerging governance of AI will indeed shape how its benefits and costs will be distributed among social groups and states.[80] As with previous technological innovations,[81] the governance of AI may result in collective benefits, or conversely favor certain actors at the expense of others.[82] The need for a global strategy stems from the global nature of these externalities, particularly in the fields of science and innovation. Multilateral institutions must facilitate knowledge sharing, the pooling of resources and the coordination of research efforts, while ensuring that progress benefits all parts of the world in an equitable manner.

76. L. Eliott, "Big Tech Firms Recklessly Pursuing Profits from AI, Says UN Head", *The Guardian*, January 17, 2024.
77. M. Schaake, "It's Already Time to Think About an AI Tax", *Financial Times*, January 8, 2024.
78. I. Bremmer and M. Suleyman, "The AI Power Paradox: Can States Learn to Govern Artificial Intelligence Before It's Too Late?", *Foreign Affairs*, August 16, 2023.
79. M. Suleyman, "Containment for AI: How to Adapt a Cold War Strategy to a New Threat", *Foreign Affairs*, January 23, 2024.
80. R. Gilpin, *The Political Economy of International Relations*, Princeton: Princeton University Press, 1987; A. Dreher and V. Lang, "The Political Economy of International Organizations", in R. Congleton, B. Grofman and S. Voigt (eds.), *The Oxford Handbook of Public Choice*, Oxford: Oxford University Press, 2019.
81. D. Drezner, "Technological Change and International Relations", *International Relations*, Vol. 33, No. 2, 2019.
82. A. Dafoe, *AI Governance: A Research Agenda*, Governance of AI Program, Future of Humanity Institute, University of Oxford, 2018.

Mitigating the consequences on other global issues

The third area of concern underlying the call for global AI governance concerns AI's impact on other global issues. Indeed, many governments and public bodies have already integrated AI into their daily activities, in order to more efficiently evaluate eligibility for social assistance, report potential fraud, profile suspects, assess risks and engage in surveillance.[83]

AI systems are not infallible, however: in reality, they frequently make mistakes, with sometimes dramatic unintended consequences. Dutch authorities recently implemented an algorithm that drove tens of thousands of families into poverty after mistakenly ordering them to repay child benefits, ultimately forcing the governor to resign.[84] In Australia, the Robodebt system, designed to detect errors in social security payments, mistakenly issued 400,000 such debts, which the Australian government had to cancel.[85]

It is estimated that no less than 85% of all projects involving AI will contain errors caused by biases of the algorithms, their developers or the data used in their operation.[86] The AI Incident Database records incidents caused by this type of error and the damage they have caused.[87] This then raises the question of how those who have suffered damages as a result of an AI malfunction should be compensated.

It is worth bearing in mind that AI and its various applications – from marketing to healthcare and weapons systems – stand to profoundly transform society and the world as a whole. It will therefore be all the more difficult to regulate this multifaceted technology, as its effects extend into such diverse areas. As the adoption of AI has the potential to impact the entire global economy, regulatory efforts must therefore reach beyond technology-specific issues.

At the same time, the global development of AI is introducing new sources of competition and tension. These technologies could exacerbate economic inequalities both within and between states, posing threats to international security. Poorly regulated applications risk disrupting nuclear stability, facilitating the development of biological and chemical weapons,

83. G. Misuraca and C. van Noordt, "Artificial Intelligence for the Public Sector: Results of Landscaping the Use of AI in Government Across the European Union", *Government Information Quarterly*, Vol. 39, No. 3, 2022.
84. Mr. Heikkilä, "AI: Decoded: A Dutch Algorithm Scandal Serves a Warning to Europe – The AI Act Won't Save Us", Politico, March 30, 2022.
85. L. Henriques-Gomes, "Robodebt: Government Admits It Will Be Forced to Refund $550m under Botched Scheme", *The Guardian*, March 26, 2020.
86. "Gartner Says Nearly Half of CIOs Are Planning to Deploy Artificial Intelligence", *Gartner Newsroom*, February 13, 2018, available at: www.gartner.com.
87. Available at: https://incidentdatabase.ai.

or democratizing the use of autonomous weapons systems. Global governance exists precisely to prevent hostility between nations from escalating into open conflict.

In order to address these issues, the design, dissemination and use of AI technologies must be regulated, to protect our collective interests. As do other general-purpose technologies – like the steam engine, electricity, or the internet[88] – AI profoundly influences economic competitiveness, military security, and our individual integrity, with consequences for states and societies.[89]

The existing and emerging governance regimes must therefore implement a wide range of policies, in accordance with decisions made regarding the oversight of AI. The international trade regime could, for example, be revised to reflect changes related to AI and its specificities.[90] Similarly, the International Fund for Cultural Diversity (IFCD), part of the United Nations Educational, Scientific and Cultural Organization (UNESCO), could have a role to play in defending linguistic diversity, which is greatly undermined by the predominance of English in LLMs. More generally, debates surrounding AI intersect with issues such as equitable economic development and human rights, which already attract the attention and investment from the international community.

The challenges facing most countries in such a wide array of issues require a coordinated response on a global scale. Global governance also plays a fundamental role in strengthening cultural exchanges and mutual understanding between nations. Institutions like UNESCO therefore help to build bridges between cultures, promote dialogue and nurture a sense of belonging within a global community.

Effective AI governance cannot, however, be limited to national or regional frameworks. Governments must work together to establish globally coordinated and interoperable standards, based on a rigorous understanding of AI-related incidents and threats. This includes the ethical oversight of data usage, the management of environmental impacts and the prevention of abuses related to algorithms. Such an approach would ensure that AI benefits society as a whole while minimizing its negative externalities. Faced with such daunting challenges, AI stakeholders and regulators have swung into action, but are moving in different directions.

88. C. B. Frey, *The Technology Trap: Capital, Labor and Power in the Age of Automation*, Princeton: Princeton University Press, 2019.
89. J. Tallberg et al., "The Global Governance of Artificial Intelligence: Next Steps for Empirical and Normative Research", *International Studies Review*, Vol. 25, No. 3, 2023.
90. E. Erman and M. Furendal, "The Global Governance of Artificial Intelligence: Some Normative Concerns", *Moral Philosophy & Politics*, Vol. 9, No. 2, 2022.

The Balkanization of governance

During the G7 summit held from June 13 to 15, 2024 in the Puglia region of Italy, Pope Francis urged policymakers to "adopt concrete actions to govern ongoing technological development [in the field of AI] towards universal fraternity and peace".[91] While there seems to be a general consensus on the need to develop clear standards for the lawful and ethical use of data, the protection of intellectual property, and limiting the environmental damage caused by AI, it is unclear what this will mean in practice.

First set in motion in 2019,[92] efforts to establish a framework for AI ramped up in the spring of 2021, when the European Commission presented the initial plan for its AI Act. China and the United States have also adopted new regulatory frameworks for AI, while numerous multilateral and civil society initiatives have emerged.

The current regulatory landscape is therefore quite fragmented.[93] Countless bodies debate the drafting of new standards despite limited diplomatic resources, any one state being unable to invest in all these forums simultaneously. This state of affairs has led to what is known as 'forum shopping', whereby states strategically select arenas with higher or lower standards of oversight or restrictions, depending on their national interests and the constraints they wish to impose on their competitors.[94] What results is a regulatory cacophony that struggles to harmonize.

This second part therefore describes the current state of international IA regulation: where and how are new regulatory measures emerging globally?

91. L. Besmond de Senneville, "Au G7, le pape François en défenseur du 'contrôle humain' face à l'IA", *La Croix*, June 14, 2024.

92. Though some early initiatives were launched by private actors beginning in 2017, notably with the AI for Good Global Summit, which grew out of a partnership between the Canadian X Prize Foundation and the International Telecommunication Union.

93. L. Schmitt, "Mapping Global AI Governance: A Nascent Regime in a Fragmented Landscape", *AI and Ethics*, Vol. 2, 2022.

94. W. Hofmann-Riem, "Artificial Intelligence as a Challenge for Law and Regulation", in T. Wischmeyer and T. Rademacher (eds.), *Regulating Artificial Intelligence*, Cham: Springer, 2020.

The approaches of the three AI "blocs"

The international AI race is centered around three main "blocs": the United States, China and Europe.[95] In this global contest, governance is just as crucial a factor as investment, data, algorithms, computing power and talent. The objective is to acquire a regulatory arsenal to support innovation.

In this area, the EU, recognized as one of the major normative powers, has managed to distinguish itself, even if the global innovation landscape remains dominated by the Sino-American duopoly. How do the approaches of these three "blocs" then differ?

The European AI Act: towards a new "Brussels effect"?

The EU has proven to be a pioneer in the regulation of AI. In 2020, the Commission published its White Paper on Artificial Intelligence, which led to debates and, on April 21, 2021, to an official proposal for European legislation on the subject. On December 6, 2022, the European Council adopted a general approach and then began negotiations with the Parliament, before reaching an agreement one year later, on December 3, 2023. On March 13, 2024, the draft regulation was adopted by the ninth legislature of the European Parliament, with 523 votes in favor and 46 against. Parallel to this, on May 17, 2024, the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law was also adopted.[96] Then, on May 21, 2024, the text was officially adopted by the 27 national ministers assembled in the Council of the EU. Published in the following weeks in the Official Journal of the EU, the AI Act came into force on August 1, 2024, to be gradually implemented by the AI Office – specially created for this purpose in January 2024 – between February 2025 and August 2027.[97] On November 14, 2024, the Commission published the first draft of the General-Purpose AI Code of

95. B. Pajot, "Artificial Intelligence, or the Race for Power", op. cit.

96. This makes it the first legally binding international instrument, opened for signature on September 5, 2024. So far, it has been signed by Andorra, Georgia, Iceland, Montenegro, Norway, Moldova, the United Kingdom, San Marino, the United States, Israel and the European Union.

97. In February 2025, Chapter I, covering general provisions, and Chapter II, covering prohibited AI practices, i.e., AI applications with "unacceptable" risk, will come into force. In August 2025, Chapter III Section 4 (Notifying Authorities and Notified Bodies), Chapter V (General-Purpose AI Models), Chapter VII (Governance), Chapter XII (Penalties) containing Article 78 (Confidentiality) will come into effect, with the exception of Article 101 (Fines for Providers) In August 2026, the act will be in force with the exception of Article 6, paragraph 1 of Chapter III and the obligations applicable to the "high" risk AI system categories. The entire act will come into force in August 2027.

Practice, which aims to facilitate the proper implementation of the regulations defined by the AI Act.[98]

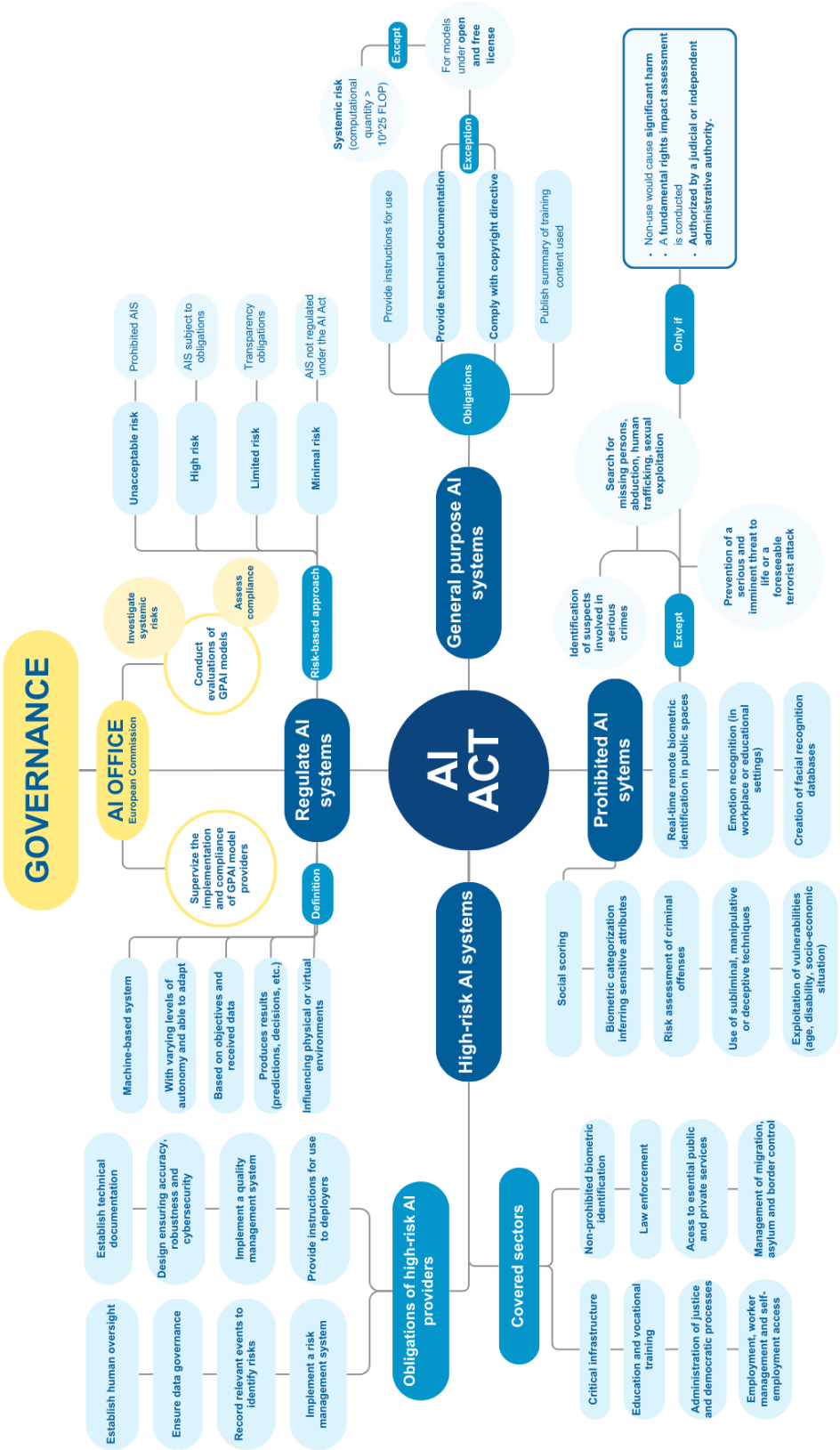
The first binding regulatory framework covering the issue, it states that AI systems and their many applications must be examined to determine what risks they may present for users; this will determine providers' obligations. While applications and systems presenting an "unacceptable"[99] level of risk, such as the social rating programs implemented by the Chinese government, are prohibited, "high-risk applications"[100] such as automatic resume screening procedures for job applicants are subject to specific legal requirements – namely in terms of quality, transparency, human oversight, data governance and security – and must be evaluated before they come to market and throughout their lifecycle.

Another, less critical section deals with "limited risk" AI systems, which are subject to more flexible transparency obligations: developers and deployers only need to inform users that they are interacting with AI (e.g., when using chatbots or deepfakes). General-purpose AI models, which can pose a systemic risk, fall into a separate category, added in 2023 due to the success of general-purpose AI systems such as ChatGPT: these are subject to additional obligations and a specific safety assessment, in particular concerning the associated systemic risks.

Uses which are not explicitly prohibited or listed as presenting a high risk largely escape regulation, however. The AI Act also does not address military applications of AI, as the European system delegates this policy area to national governments. This categorization of risks and its resulting obligations must nevertheless ensure that the AI systems used in the EU are safe, transparent, traceable, non-discriminatory and environmentally friendly.

98. As proposed by independent experts appointed by the AI Office, this code's main features include details on transparency and the application of copyright rules for providers of general-purpose AI models, as well as a taxonomy of systemic risks, risk assessment methods and mitigation measures for providers of advanced general-purpose AI models that may present systemic risks.
99. This label pertains to AI applications that are deemed to undermine European values and fundamental rights, such as human dignity, democracy or the rule of law.
100. This includes AI systems used in healthcare, education, recruitment, critical infrastructure management, policing and the justice system.

The key elements of the AI Act



This risk-based approach could serve as a model for the development of global standards, based on transparency, security, responsibility and the protection of human rights. The EU's stated objective here is to trigger a new 'Brussels effect' with the first major AI regulation, promoting European standards and values around the globe.[101] Such was the case in 2019, when the General Data Protection Regulation (GDPR) was enacted: shortly after its ratification, the Data Free Flow with Trust initiative, itself derived from the GDPR, was approved by the G20 countries in Osaka (Japan); various other measures directly inspired by the European regulation have since followed in Brazil, China, South Korea and Chile.

As it is thus far the most successful regulatory initiative in this area, the AI Act could in fact serve as a starting point for multilateral talks and become a model for national legislation. Given the considerable influence it commands in international negotiations because of its market share, the EU is also in a strong position to persuade other countries to adopt binding regulations on AI. As Jean-Rémi de Maistre has said, "the AI Act could prove to be an important strategic asset for Europe".[102] Like the GDPR, it applies extraterritorially to goods and services used and circulating in the European market but having an impact on foreign providers: it thus extends to any entity providing an AI-based product or service in the Union.

Its implementation across the EU faces a number of challenges, however, ranging from harmonization between EU Member States to stakeholder involvement, including government entities and suppliers, importers, users and distributors of AI systems. The impact of this regulation will also depend on how other regions respond, and in particular the major AI powers like the United States and China.[103] If no global alignment is achieved, this could result in a patchwork of regulations, undermining compliance.

The AI Act negotiations were also not well received by certain prominent start-ups in the field, chief among them the French "unicorn" Mistral AI, which reached a valuation of some 6 billion euros in just a few months. As its founder Arthur Mensch argued on X in November 2023, the European regulation would favor "established businesses that can afford to meet heavy compliance requirements", namely the digital giants and their "army of lawyers".[104]

101. A. Bradford, *The Brussels Effect: How the European Union Rules the World*, Oxford: Oxford University Press, 2020.
102. J.-R. de Maistre, "Intelligence artificielle : 'L'innovation ne doit pas se faire au sacrifice de l'éthique et de la sécurité'", *Le Monde*, September 11, 2024.
103. R. Csernatoni, "The AI Governance Arms Race: From Summit Pageantry to Progress?", Carnegie Endowment for International Peace, October 7, 2024.
104. C. Auffray, "Pour Arthur Mensch (Mistral AI), l'*AI Act* se trompe de cible", *ZDNet*, May 22, 2024.

China vying for leadership in global AI governance

While the AI Act was being negotiated in Europe, the People's Republic of China (PRC) was itself beginning to reflect on how to develop its approach to regulating the sector. This initiative fits into a broader trend in China, which aspires to overhaul international governance, as it believes its interests are not properly taken into account in the current system.[105] Its objective is also, simultaneously, to exercise strict control of digital technology at the national level.

In 2021 and again in 2022, China became the first country to adopt detailed, strict regulations governing some of the most widespread applications of AI, which laid the foundations for a new governance framework. This evolving political structure will have major geopolitical repercussions, as it impacts exploratory AI research, the running of the world's second-largest economy, and issues as wide-ranging as large language models in Africa and autonomous vehicles in Europe.[106] On July 2, 2024, Beijing again announced that it intends to enact more than fifty new standards concerning AI by 2026.[107]

Since 2017, China has sought to establish itself as the global leader in AI by 2030. This strategic objective is based on an ambitious development plan for next-generation AI, and in particular, generative AI. This project comes with massive funding: in May 2024, Reuters reported that the Big Fund, an investment fund specializing in semiconductors and backed by the Chinese government, was entering its third investment phase, with a budget of 344 billion yuan, or approximately 44.12 billion euros.[108]

In terms of standards, Chinese policymakers have also stressed their determination to act first, in order to secure global leadership in AI governance. Since 2021, several pieces of legislation have been introduced to regulate Chinese innovation. However, unlike the EU, which has adopted a holistic approach based on risk mitigation and upholding values, China sequentially focuses on specific applications of AI through different sets of rules imposing new obligations, first in the field of algorithmic recommendations, then with regard to the techniques used to generate synthetic content (deepfakes).[109]

105. S. Kastner, M. Pearson and C. Rector, *China's Strategic Multilateralism: Investing in Global Governance*, Cambridge: Cambridge University Press, 2019.
106. M. Sheehan, "Tracing the Roots of China's AI Regulations", Carnegie Endowment for International Peace, February 2024.
107. J.-R. de Maistre, "Intelligence artificielle : 'L'innovation ne doit pas se faire au sacrifice de l'éthique et de la sécurité'", op. cit.
108. "China Sets Up Third Fund with $47.5 Bln to Boost Semiconductor Sector", Reuters, May 27, 2024.
109. Mr. Sheehan, "Tracing the Roots of China's AI Regulations", op. cit.

Under these rules, providers are required to identify AI-generated content and ensure that it does not violate people's image rights or damage the nation's image. These two regulations have also led to the creation of an algorithm registry, which has become the cornerstone of China's AI governance regime.

In July 2023, China also took a major step in the regulation of generative AI, adopting a regulatory framework aimed specifically at these services.[110] Its aim is to balance technological progress and safety requirements, while promoting the growth of the industrial sector. These measures apply to publicly available generative AI services in China, regardless of whether the provider is based in China or abroad. The measures relating to generative AI therefore have an extraterritorial effect, as with the AI Act.

As researcher Matt Sheehan argues, contrary to popular belief, China's AI governance regime was not developed through a top-down process. While President Xi Jinping and other senior Chinese Communist Party (CCP) officials may set policy priorities, they are not the main drivers of China's AI regulations. They are actually developed through a dynamic and iterative process involving multiple actors both internal and external to the Chinese government. These actors include middle-level bureaucrats, academics, tech representatives, journalists and researchers from companies in the digital sector. Through a mix of public advocacy, intellectual debate, technical workshops, and bureaucratic wrangling, these actors laid the foundation for China's current and future AI regulations.[111] These normative efforts also appear to result mainly from fears of AI-driven political destabilization, which has prompted the CCP to tightly regulate the private sector.

China seems to share Western concerns regarding the risks posed by AI.[112] A growing number of research papers, public statements and government documents suggest that AI security is becoming a subject of increasing importance in China, prompting both massive technical investment and regulatory action. Interest in this topic first grew within the Chinese technocracy, before spreading to the country's highest political spheres. In July 2024, while Shanghai was hosting a global conference on AI, the CCP published a policy document calling for the creation of "oversight systems to ensure the safety of artificial intelligence".

On the international stage, China makes its presence felt in its efforts to influence the standardization of AI on several levels. Between 2014 and 2023, it filed over 60% of all patents in generative AI worldwide, three times more

110. M. Rochefort, "La Chine s'apprête à réglementer l'IA générative", *Siècle Digital*, July 11, 2023.
111. Mr. Sheehan, "Tracing the Roots of China's AI Regulations", op. cit.
112. Mr. Sheehan, "China's Views on AI Safety Are Changing – Quickly", Carnegie Endowment for International Peace, August 2024.

than the United States, though they struggle to establish themselves as global leaders.[113] Beijing also stands out through its active participation in international standardization bodies, in particular the International Telecommunication Union. China also launched its own Global AI Governance Initiative in October 2023 and is actively working on AI capacity building, particularly within BRICS+.

Given the United States' disengagement from multilateral affairs following Donald Trump's inauguration, China may be tempted to reclaim the West's narrative for its own purposes. It could continue its capacity-building efforts in the Global South, including technology transfers and access to emerging markets. Simultaneously, it could promote its own technological standards more forcefully in order to expand its influence and capture more markets.

Regulations undermined in the United States

While Europe and China pursue their strategies, the United States has devised its own roadmaps to identify and mitigate the threats posed by AI technologies. While the U.S. narrative of a "zero-sum game" with China limits its ability to pursue ambitious regulations, internal power struggles have led to a phenomenon of legislative inflation. These initiatives seek to address voters' concerns while reasserting government institutions' regulatory authority.[114] U.S. Senator Chuck Schumer declared in April 2023 that it was important for the United States not to let China "lead on innovation or write the rules of the road"[115] for AI.

The Biden administration's approach to AI governance first materialized as the Blueprint for an AI Bill of Rights, published in October 2022. This framework sets out five key principles to ensure the responsible development of these technologies, including safeguards against algorithmic bias and privacy considerations. It has not been adopted to date, however, and serves only as a non-binding guide for agencies and businesses involved in the development or deployment of AI systems.

In collaboration with the private and public sectors, the National Institute of Standards and Technology (NIST) then developed the AI Risk Management Framework to better address the risks posed by AI – to individuals, organizations and society. Intended for voluntary use, it aims to improve the ability to incorporate considerations of trustworthiness into the design,

113. "Generative Artificial Intelligence", *Patent Landscape Report*, World Intellectual Property Organization, July 2024.
114. R. Heath, "Exclusive: States Are Introducing 50 AI-related Bills per Week", *Axios*, February 14, 2024, available at: www.axios.com.
115. D. Shepardson, "US Senate Leader Schumer Calls for AI Rules as ChatGPT Surges in Popularity", Reuters, April 13, 2023.

development, use and evaluation of AI products, services and systems in the United States.

In September 2023, the Biden-Harris administration also secured voluntary commitments from industry-leading companies for the safe and transparent development of AI technologies.[116] These commitments include ensuring that products are safe before being released to the public, and establishing systems focused on security and public trust.

One month later, Washington released its first major decree on AI, President Biden's Executive Order 14110, establishing an array of standards, security measures, privacy protections and controls on the "safe, secure and trustworthy development and use of artificial intelligence" This decree sought in particular to address issues of fairness and civil rights, while also touching on specific applications of AI. It did so with robust, reliable, reproducible and standardized evaluations of AI systems, as well as the establishment of public policies, institutions and mechanisms to test, identify and mitigate the risks associated with these systems before they are put into use. Just as with China and the EU, Executive Order 14110 contained clauses with extraterritorial reach.[117]

This executive order was a crucial step in the definition of regulations and responsibility in the field of AI in multiple sectors, focusing on a "whole-of-government" approach to address both the opportunities and risks associated with AI. It was one of the U.S. government's most ambitious efforts to regulate the development and use of this technology, in an attempt to establish the United States as a leader in the adoption of safe, ethical and responsible AI practices. The Executive Order tasked federal agencies with several major challenges: managing dual-use AI models, establishing rigorous testing protocols for high-risk systems, implementing accountability mechanisms, defending civil rights and ensuring transparency at every stage of the AI lifecycle. These measures aimed to reduce security risks, protect democratic values and strengthen public confidence in this constantly evolving sector.

But in the run-up to Trump's second term in office, the fate of this initiative seemed uncertain. After announcing in July 2024 that he would revoke the order if elected,[118] the president fulfilled his promise only a few

116. "Fact-Sheet: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage Risks Posed by AI", Washington, D.C.: White House, July 21, 2023, available at: www.whitehouse.gov.

117. Specifically, a clause that requires the Department of Commerce to "propose regulations that require United States internet Infrastructure as a Service providers to submit a report to the Secretary of Commerce when a foreign person transacts to train a large AI model with potential capabilities that could be used in malicious cyber-enabled activities".

118. Beyond the federal level, local initiatives are emerging, such as two bills introduced in California to regulate the use of AI in employment.

hours after his inauguration on January 20, 2025.[119] The following day, he announced the launch of the Stargate project, with a budget of $500 billion over four years to build "the physical and virtual infrastructure to power the next generation of AI".[120]

As much of the AI regulation effort relies on the work of federal agencies, the principles established in the decree could endure even after it is repealed, allowing the United States to preserve its influence in AI standard setting. The increasingly prominent role played by American tech magnates in the new administration, however – including investor David Sacks, appointed White House AI & Crypto Czar[121] – undermines efforts to regulate AI, as they champion a decidedly anti-regulation agenda. In any case, the long-term impact of the Biden–Harris administration's efforts will depend on the ability of policymakers to keep pace with the rapid advances in AI, while maintaining a balance between supporting innovation and public trust.[122]

A patchwork of multilateral initiatives

The race for technological innovation brings with it fierce competition between countries, many of which dream of becoming the leading standard-setting power for AI. The three major blocs have therefore sought to take the lead in governance, in order to establish regulations that are compatible with their national ambitions, while also potentially stalling their competitors. But this patchwork of regulations raises fears of a Balkanization of governance, with competing or even contradictory standards being adopted, fragmenting governance and undermining the ambition of universal regulation for an intrinsically universal technology.

Thus have a multitude of multilateral alternatives emerged in recent years. Firstly, the Global Partnership on AI (GPAI), established in 2019 and first proposed by Canada and France, brings together 29 members (28 states and the EU) with the aim of strengthening North-South multi-stakeholder cooperation to foster the development of AIs that reflect democratic values and can respond to global challenges. The GPAI also advocates for responsible and human-centered use of AI, in accordance with human rights and fundamental freedoms. Hosted by the Organization for Economic Co-operation and Development (OECD), it also relies on two AI research centers, the French National Institute for Research in Digital Science and

119. D. Shepardson, "Trump Revokes Biden Executive Order on Addressing AI Risks", Reuters, January 21, 2025.
120. S. Holland, "Trump Announces Private-sector $500 Billion Investment in AI Infrastructure", Reuters, January 22, 2025.
121. A. Chow, "What Trump's New AI and Crypto Czar David Sacks Means For the Tech Industry", *Time*, December 10, 2024.
122. A. Kein, C. Kerry, C. Radsch, M. MacCarthy, S. Friedley and N. Turner Lee, "1 Year Later, How Has the White House AI Executive Order Delivered on Its Promises?", Brookings Institution, November 4, 2024.

Technology (INRIA) and the Canadian Center of Expertise in Artificial Intelligence (CEIMIA) in Montreal, to promote collaboration and knowledge sharing between civil society, governments and academia. The GPIA's work is organized around four expert working groups: Responsible AI, Data Governance, Future of Work, and Innovation and Commercialization. Due to its members' limited geographical distribution and the lack of binding agreements, the GPAI is regularly criticized for its lack of universality, representativeness – with only seven countries from the Global South, in addition to the OECD and twelve EU countries – and effectiveness.

GPIA member countries

In 2019, the OECD developed a set of AI principles.[123] Adopted by 47 countries, they promote innovative and trustworthy AI that respects human rights and democratic values. Their update in 2024, taking into account new technological and political developments, also demonstrates the need for "future-proof" regulation to ensure that established standards remain robust and fit for purpose even as technologies evolve and new applications emerge.

123. More specifically, they consist of five principles based on values and five recommendations intended to serve as practical and flexible guidelines for policymakers and AI stakeholders.

Countries adhering to the OECD AI principles

The OECD principles have a significant influence: Member States draw on these guidelines to establish their regulatory frameworks for AI, thereby laying the foundations for global interoperability between their respective jurisdictions. This is the case, for example, for the EU, the Council of Europe, the United States and even the United Nations, which use the OECD definition of an AI system in their legislative and regulatory frameworks and policy guidelines.

Comparative table of UNESCO and OECD principles on artificial intelligence

Principles and recommendations	UNESCO	OECD
Fairness, human values and non-discrimination	✓	✗
Transparency and explainability	✓	✗
Security, safety and robustness	✓	✗
Accountability	✓	✗
International cooperation and AI governance	✓	✗
Proportionality and harmlessness	✓	✗
Right to privacy and data protection	✓	✗
Sustainability	✓	✗
Awareness and education	✓	✗
Monitoring and human decisions	✓	✗
Inclusive growth, sustainable development and well-being	✓	✗
Investing in AI R&D	✓	✗
Shaping an AI-friendly policy framework	✓	✗
Fostering a digital ecosystem for AI	✓	✗
Strengthening human skills and preparing for the transformation of the labour market	✓	✗

In **bold**, OECD recommendations intended for policymakers.

In November 2021, UNESCO proposed its first global standard on AI, the "Recommendation on the Ethics of Artificial Intelligence", which was adopted by all 193 of its Member States. Central to this recommendation is the protection of human dignity and human rights through the reinforcement of fundamental principles such as transparency, fairness, accountability and oversight of AI systems. UNESCO has also included many areas for strategic action, aimed at helping policymakers translate values into action, particularly with regard to data governance, the environment, gender, education, research, health and social welfare. Though non-binding, this recommendation found some success in

February 2024, when eight major tech companies, including industry giant Microsoft, officially committed to upholding its values and principles at every stage of the design and deployment of their AI systems.[124]

During the G7 in May 2023, under Japan's presidency, the Hiroshima Process was also launched, aimed at defining the main principles of generative AI governance and, more broadly, advanced AI models ("frontier AI"). Led by Japan, this initiative is made up of several working groups created specifically to address these issues. Participants include France, Canada, the United States, the United Kingdom, Italy, Germany and the European Commission, with the OECD and the GPIA contributing as guest organizations.

In October 2023, the G7 countries also agreed on a voluntary code of conduct for companies developing advanced AI systems. The eleven points of this code aim to "promote safe, secure and trustworthy AI worldwide" and to "help seize the benefits and address the risks and challenges brought by these technologies". Companies are encouraged to take appropriate measures to identify, assess and reduce risks throughout the lifecycle of AI systems, and to address any incidents involving AI products already on the market. The code also calls on them to implement robust security controls and to be transparent about their capabilities and any obstacles they encounter. The G20, too, plays its part in this wide-ranging drive to regulate AI. During Brazil's presidency in 2024, the focus was placed on addressing AI from the perspective of economic development and developing countries' access to AI and its infrastructures.

Another trend in multilateral efforts to regulate AI technologies is the emergence of a form of "summit diplomacy", in which different states seek to assert themselves as leaders in global governance by organizing large international conferences. At the behest of Prime Minister Rishi Sunak, in the hopes of establishing it as his political legacy, the United Kingdom organized the first ever global summit on AI, the Artificial Intelligence Safety Summit, in November 2023 at Bletchley Park. This event, which brought together multilateral actors and a wide range of stakeholders to discuss the risks associated with generative AI, was not only a diplomatic success for the UK, but also initiated an encouraging dynamic of cooperation, even if the role played by major private sector participants in the discussions was criticized.[125]

Prior to the summit, Rishi Sunak had also announced the creation of an AI Safety Institute, designed to be replicated by the other states participating in the summit, as the United States has done within NIST. On

124. P. Rioux, "Intelligence artificielle : 8 géants de la tech s'engagent à appliquer la recommandation éthique de l'Unesco", *La Dépêche*, February 7, 2024.
125. M.-F. Cuéllar, "The UK AI Safety Summit Opened a New Chapter in AI Diplomacy", Carnegie Endowment for International Peace, November 9, 2023.

November 20 and 21, 2024, the first AI Safety Institutes and the government-mandated offices in Australia, Canada, the European Commission, France, Japan, Kenya, the Republic of Korea, Singapore, the United Kingdom and the United States met in San Francisco for the first meeting of the International Network of AI Safety Institutes.[126] These institutes could move beyond their original mission and become a source of international standards, in a more bottom-up approach, focusing on concrete technical issues.[127]

At the end of this first AI Safety Summit, 28 states including China, the United States and the EU also agreed on a joint declaration, known as the "Bletchley Declaration", expressing their desire to cooperate in the establishment of a regulatory framework to ensure the responsible and trustworthy development and use of AI worldwide. After the United Kingdom, South Korea and France organized their own global AI summits. The Seoul summit, held in May 2024, once again focused on security, and resulted in sixteen major companies in the industry (in China, South Korea, the United Arab Emirates and the United States) pledging to work on risk mitigation.[128]

To stand out from the previous summits, Paris took a resolutely positive and optimistic approach to assert itself as one of the world capitals of AI. France therefore decided to focus on action, the theme of the summit held on February 10 and 11, 2025, to showcase the national AI ecosystem and its strategy in tackling the challenges of AI. "We will highlight the risks posed by AI, which have already received considerable attention in London in November 2023 and in Seoul in May, as well as the opportunities and advantages offered by this technology", the Élysée Palace explained.[129] France also focused on further involving civil society, including a broader range of issues,[130] and trying to reconcile security concerns with support for entrepreneurship. This event could also open the way for institutional innovations, such as the creation of a "World AI Organization"[131] or a "World Data Organization",[132] to develop collective solutions to these global

126. As part of the Trade and Technology Council established by Joe Biden and Ursula von der Leyen between 2021 and 2024, the AI Safety Institutes have been called upon to extend cooperation for safe, reliable, and responsible AI.
127. D. Milmo, "UK's AI Safety Institute 'Needs to Set Standards Rather Than Do Testing'", *The Guardian*, February 11, 2024.
128. Among them are Amazon, Google, Microsoft, Meta, Mistral AI and OpenAI. See "Seoul Declaration for Safe Innovative and Inclusive AI: AI Seoul Summit 2024", Department for Science, Innovation and Technology, available at: www.gov.uk.
129. A. Piquard, "Le sommet de Paris vise à agir contre les risques et surtout pour les bénéfices de l'IA", *Le Monde*, December 9, 2024.
130. The summit will focus on five areas: public interest, jobs, culture, ethics and global AI governance.
131. As suggested in the report of the National AI Commission. See "AI: our ambition for France", French Artificial Intelligence Commission, March 2024.
132. I. Bremmer, "Why We Need a World Data Organization. Now", *GZERO*, November 25, 2019.

challenges and establish a framework for constructive cooperation around these issues.

The countries of the Global South seem to be following their Western counterparts' lead: Rwanda has announced that it intends to host the first international summit on AI in Africa.[133] India is also heavily invested in both the development and governance of AI,[134] as evidenced by its GPIA presidency in 2024 and its co-presidency of the AI Action Summit in February 2025. However, this focus on international summits may have some pernicious effects, as researcher Stuart Russell observed in July 2024: "the whole mindset has become one of AI as a vehicle for economic nationalism[135]", and a tool for nation branding.

The UN would appear best placed, in this respect, to prevent forum shopping, and to develop robust and universal regulations for these technologies. Steps are gradually being taken in this direction: on July 18, 2023, the United Nations Security Council held its first meeting dedicated to AI. "The very nature of the technology itself – transboundary in structure and application – necessitates a global approach", the final report by the group of experts appointed by Antonio Guterres to examine the issue concluded.[136] In addition, in March 2024, the United Nations General Assembly adopted – with 193 votes in favor – a resolution aimed at establishing international rules governing the use of AI, to "bridge the digital divide" and mitigate risks.[137] What concrete measures will result from this report and resolution remains unclear, however, all the more so given that the expert group's findings are advisory and the resolution is not legally binding.

The slow progress of the GPAI – and, more recently, of the summits on AI security – also demonstrates the difficulty of establishing global regulations in a deeply fragmented world. It is not enough for States to join a new

133. "Rwanda Announces Plans to Host Inaugural High-Level Summit on AI in Africa", Rwanda Centre for the Fourth Industrial Revolution, January 18, 2024.
134. In March 2024, the country adopted an AI strategy (IndiaAI Mission), with a budget of 1.2 billion dollars over five years to be distributed among seven pillars, including governance. The government wants to promote safe and trusted AI, using applications developed by public authorities as models for good conduct and developing technical tools to ensure AI applications conform to a certain ethical framework.
135. A. Piquard, "À l'approche du sommet de Paris, les militants inquiets quant à la 'sécurité de l'IA' cherchent à se faire entendre", *Le Monde*, September 11, 2024.
136. The United Nations High-Level Advisory Board on AI, established on October 26, 2023, is made up of 39 experts from around the world. See "'Irrefutable' need for global regulation of AI: UN experts", *UN News*, September 19, 2024, available at: news.un.org.
137. This resolution excludes defense. In general, the issue of military AI governance focuses primarily on the subject of autonomous weapons and is treated as a separate international debate, as part of the Convention on Certain Conventional Weapons (CCW) in Geneva. However, while the group of government experts mandated to discuss the normative and operational framework for these technologies has been working for more than ten years, regulatory efforts have reached an impasse. For an in-depth analysis of this regulatory initiative, see L. de Roucy-Rochegonde, *La Guerre à l'ère de l'Intelligence artificielle*, Paris: PUF, 2024.

international institution for standards to materialize, all the more so in the field of AI, where non-State actors play an essential role.

Alternatives from non-state actors

Efforts to establish international governance for AI would fall short without the active participation of those developing the technology, in particular the U.S.'s leading digital platforms. As the Secretary-General of the United Nations, Antonio Guterres, said in Davos in January 2024: "We need governments urgently to work with tech companies on risk management frameworks for current AI development, and on monitoring and mitigating future harms". These companies do not, however, tend to be entirely cooperative, as they often regard regulation as an obstacle to innovation. Their lobbying efforts in international negotiating forums are such that they can now influence the broad outlines of national and international AI governance.[138]

The U.S. digital giants have thus adopted an aggressive approach to influence policy decisions in the White House, the administration, Congress, and even in Europe. Their aim is to steer regulation towards long-term risks, and away from short-term concerns, in order to afford themselves greater freedom of action.[139] This strategy reveals a striking contradiction on the part of such prominent figures as Elon Musk and Sam Altman, who, despite their alarmist discourse on artificial general intelligence (AGI) and their claimed advocacy of strict regulation, play an active part in this technological arms race while seeking to undermine regulatory frameworks, in effect deflecting their own responsibility onto regulators. Even when they claim to fear outright catastrophe, as they did in an open letter published in May 2023 calling for a moratorium on generative AI, the aim is only to hold their competitors back while they make up for their technological shortcomings.

As researcher Courtney Radsch warned in November 2024, neither ambitious measures like the European AI Act nor more nuanced initiatives like Joe Biden's Executive Order are likely to "mitigate the monopolistic power of a handful of tech giants".[140] At a time when multiple Silicon Valley executives have pledged their allegiance to Donald Trump ahead of his inauguration on January 20, 2025,[141] a strongly anti-regulation and even

138. J. Tallberg et al., "AI Regulation in the European Union: Examining Non-State Actors Preferences", *Business and Politics*, Vol. 26, 2024.
139. B. Pajot, "Les risques de l'IA : enjeux discursifs d'une technologie stratégique", op. cit.
140. A. Kein, C. Kerry, C. Radsch, M. MacCarthy, S. Friedley and N. Turner Lee, "1 Year Later, How Has the White House AI Executive Order Delivered on Its Promises?", op. cit.
141. A. Leparmentier, "Après Musk et Bezos… Zuckerberg : la tech en ordre de marche derrière Trump", *Le Monde*, January 8, 2025.

anti-European agenda seems set to prevail. On January 7, 2025, Mark Zuckerberg, the CEO of Meta, Instagram and WhatsApp, announced that he would end his fact-checking partnerships with several major U.S. and international media outlets in favor of a content moderation policy modeled on the "community notes" feature already present on X. In his press release, he also had strong words for Europe, which he said "has an ever-increasing number of laws [...] making it difficult to build anything innovative there". He went on to announce his willingness to work with Donald Trump to "push back on governments around the world that are going after American companies".[142]

This support for the newly installed president seems to signal the start of an offensive against Europe, targeting the fines and taxes imposed by the European Commission – interpreted as a means of penalizing U.S. companies – as well as regulations deemed too restrictive and stifling to innovation. Though the tech giants have complied with data protection laws, they now seem determined to categorically oppose regulation on AI, which they deem excessive, an issue they see as crucial for the United States.

These constant complaints about the supposed incompatibility between innovation and regulation are nothing new from the leaders of Big Tech.[143] In September 2024, some thirty digital companies published an open letter denouncing Europe as having "become less competitive and less innovative compared to other regions, and it now risks falling further behind in the AI era due to inconsistent regulatory decision making". Mark Zuckerberg went even further, suspending the EU launch of his AI assistant, Meta AI, on Instagram and Facebook.[144] In its press release, the group urged Brussels not to "reject progress [...] and watch as the rest of the world builds on technologies that Europeans will not have access to". In reality, this debate goes beyond social, economic, and legal issues; it is part of a broader strategic vision. The narrative of technological competition serves to promote a minimalist regulatory model, which is touted as crucial to preserving innovation and maintaining a competitive advantage over international rivals.

It must be stressed, however, that the hypothesis according to which regulation stifles innovation and weakens Western states geopolitically is unproven. This idea, promoted by the major U.S. stakeholders in the sector, is often presented as a warning, also unproven, against the risk of Chinese

142. D. Leloup and A. Piquard, "La fin des partenariats de fact-checking chez Meta, un revirement symbolique", *Le Monde*, January 7, 2025.
143. On U.S. accusations directed at Europe and the importance of the innovation versus regulation debate, see M. Velliet, "Digital Sovereignty: European Policies, American Dilemmas", *Notes de l'Ifri*, Ifri, March 2023, available at: www.ifri.org.
144. A. Piquard and V. Malingre, "IA: Meta et Apple mettent la pression sur l'Union européenne, accusée de 'rejeter le progrès'", *Le Monde*, September 25, 2024.

technological superiority. The European approach specifically attempts to resolve this purported tension between innovation and regulation, and to show that the two are complementary. Indeed, by providing a clear legal framework, unifying the regulations of individual national markets at the European level, and facilitating access to these markets for other actors besides the tech giants, European standards actually appear to foster competition and innovation, while at the same time enabling the smooth development of these technologies.

To try to reassure their users without having their hands tied by legally binding regulations, many companies have approached the issue of AI oversight from an ethical perspective. In September 2016, several major U.S. tech companies formed a partnership dedicated to ethics in AI: the Partnership on Artificial Intelligence to Benefit People and Society.[145] Such caution can be explained by what they perceived as a reputational risk, fueled by NGO "name and shame" campaigns that sought to tarnish the reputation of actors engaged in allegedly harmful activities.

This practice is sometimes referred to as "ethics washing". Similar to greenwashing, it involves feigning ethical concern to improve the public perception of a person or organization. This practice is particularly noticeable in many projects focused on the development and use of new technologies. In their article "Why Are We Failing at the Ethics of AI?", Anja Kaspersen and Wendell Wallach write:

> "The last few years have seen a proliferation of initiatives on ethics in artificial intelligence (AI). Whether formal or informal, led by companies, governments, international and non-profit organizations, these initiatives have developed a plethora of principles and guidance to support the responsible use of AI systems and algorithmic technologies. Despite these efforts, few have managed to make any real impact in moderating the effects of AI."[146]

The two researchers also describe ethics washing in the field of AI as "creating a superficially reassuring but illusory sense that ethical issues are being adequately addressed, to justify pressing forward with systems," even while these systems run counter to existing safeguards. Approaching the issue from an ethical perspective also has the advantage of resulting in non-binding and highly subjective commitments, unlike regulations, which necessarily involve oversight and control measures. Moreover, in terms of democratic legitimacy, the entities responsible for adopting AI regulations,

145. Which included Google, Facebook, IBM, Microsoft and Amazon. See M. Tual, "Intelligence artificielle : les géants du Web lancent un partenariat sur l'éthique", *Le Monde*, September 28, 2016.
146. A. Kaspersen and W. Wallach, "Why Are We Failing at the Ethics of AI?", Artificial Intelligence and Equality Initiative, Carnegie Council for Ethics in International Affairs, November 10, 2021.

and the grounds on which these decision-making entities are granted such power, are of great importance.[147]

Even with such stated intentions, Big Tech's real priorities can be questioned: as AI applications continue to grow, the teams working on ethical issues and responsible AI practices at Meta, Google, Microsoft and Amazon find their numbers dwindling.[148] In May 2024, OpenAI announced that it was disbanding the team responsible for the security of a potential artificial superintelligence, and redirecting its efforts towards technological development.[149]

147. E. Erman and M. Furendal, "Artificial Intelligence and the Political Legitimacy of Global Governance", *Political Studies*, Vol. 72, No. 2, 2022.
148. G. de Winck and W. Oremus, "As AI Booms, Tech Firms Are Laying Off Their Ethicists", *The Washington Post*, March 30, 2023, available at: www.washingtonpost.com.
149. "OpenAI: dissolution de l'équipe chargée de la sécurité d'une potentielle superintelligence artificielle", *Le Monde*, May 18, 2024.

For inclusive and lasting AI governance

It therefore remains to be seen whether these myriad initiatives will result in tangible commitments and help to mitigate the most concerning risks of AI in the years to come and, above all, whether they will have any impact outside the Western world, as only seven countries participate in every initiative, and 119 participate in none. Although both public and private actors have taken steps to mitigate risks associated with AI, their approaches remain disparate. As attempts at regulation multiply around the world, their coherence must be addressed to prevent fragmentation, which could lead to tension between opposing or incompatible models. Greater international cooperation is therefore necessary to combat this fragmentation, establish robust safeguards, and build inclusive and lasting AI governance.

Inclusivity as a means of achieving consensus

It is now clear that global governance is crucial to coordinate regulations, prevent legal "grey areas" from being exploited, and avert the danger of an unethical global AI race. Multilateral forums would seem to be the ideal place to harmonize the varying approaches to regulation. That being said, building a global consensus on these issues will be no easy task. This process is, in fact, proving to be extremely delicate because it requires the formulation of diplomatic language specific to novel technical issues, because nations fear falling behind technologically and strategically, and because these debates come at a time of profound crisis for multilateralism,[150] with major international cooperation agreements coming under challenge.

The differing priorities and perspectives being advocated for concerning AI governance reflect distinct national interests. Some states seek to protect their businesses, which believe regulation compromises their competitive advantage. France's reservations during negotiations on

150. Multilateralism is currently the subject of both political and academic debate because it is both a functional instrument and a political invention. Having suffered multiple crises (of vitality, functionality, universality…), it has been criticized as an organizational model (due to the politicization and depoliticization it can cause in international relations) but also as a concept. Here, it is to be understood in the most literal sense, as a major form of international action. For a substantial analysis of these debates, see J. Fernandez and J.-V. Holeindre, *Nations désunies ? La crise du multilatéralisme dans les relations internationales*, Paris: CNRS Éditions, 2022.

the AI Act were a particularly clear illustration of this approach. While purporting to defend innovation, Paris sought in reality to facilitate the growth of the promising (and above all French) company Mistral AI, which develops open source and proprietary language models for generative AI applications.

Beyond seeking a competitive advantage, stakeholders also want to impose a value model to govern the development and deployment of AI, as exemplified by the EU, which strongly prioritizes respect for human rights. The countries of the Global South, on the other hand, are more concerned with the social consequences and economic inequalities caused by these emerging technologies. They are therefore torn between two perspectives: regulation as a necessary means of mitigating negative externalities, of which they are often the first victims; or regulation as a barrier to technological emancipation. International AI governance must therefore be backed with systematic efforts to improve access to these innovations: developing economies must benefit from their potential, in order to bridge the digital divide, rather than widening it. The Paris AI Summit aims to be a step in this direction, as new Global South countries are expected to be announced as members of the GPAI, and a foundation dedicated to lowering barriers to entry and strengthening access to the digital commons is to be established in Paris. A fund-raising campaign with a target of 2.5 billion euros has been launched to this end.[151]

More generally, the consensus-building process will only succeed if all actors can be heard. As noted in the report of the UN expert group on AI: "Equity demands that more voices play meaningful roles in decisions about how to govern technology that affects us".[152] Civil society representatives also have a role to play in these discussions. They are, in fact, perhaps best placed to counterbalance the outsized influence of the digital giants in this regulatory landscape. Examples can be found in the field of arms control, in which, despite governmental reluctance and intense lobbying by major defense companies, several major international conventions resulted from civil society efforts to halt the catastrophic humanitarian consequences of certain weapons[153].

Finally, the various initiatives that make up the IA "regime complex"[154] must be addressed in a more integrated and coherent manner. The 2024 Summit on the Future, held on September 22 and 23, 2024, in New York, was an opportunity for the UN and the OECD to announce a new

151. This is also a means of preventing a Chinese monopoly on capacity building. Author's interview with an advisor to the President of France.
152. "'Irrefutable' need for global regulation of AI: UN experts", op. cit.
153. This is referred to as humanitarian disarmament, where civil society forces its way into the halls of diplomatic negotiations. Such actions have already led to the banning of blinding laser weapons, anti-personnel mines and cluster munitions.
154. R. Csernatoni, "Charting the Geopolitics and European Governance of Artificial Intelligence", Carnegie Europe, March 6, 2024.

collaboration to strengthen the global governance of AI. In his statement, Amandeep Singh Gill, the United Nations Secretary-General's Envoy on Technology, emphasized the crucial importance of collaboration between these two entities: "the speed of AI technology development and the breadth of its impact requires diverse policy ecosystems to work more cohesively. And in real time".[155]

This new initiative will take advantage of the complementary strengths of the UN, with its universal reach, and the OECD, with its proven technical and analytical expertise, to help governments act quickly and effectively to address the full breadth of issues at stake in AI. The stated objective is to establish solid global governance mechanisms, in collaboration with major stakeholders, including leading universities and internationally renowned researchers.

In the same vein, the AI Action Summit in Paris aims to build international consensus around a common foundation for AI governance. To achieve this objective, an inclusive co-construction process is planned with more than 70 stakeholders, including states, international organizations, researchers, businesses, and NGOs. The aim is to bring together representatives from these different fields and establish extensive consultation mechanisms, in order to arrive at a shared vision and positions that best reflect collective preferences and interests.

What kind of regulatory body?

The objective, however, is not just to achieve the joint adoption of regulations, but also to guarantee their effective application at the international level. Multilateral institutions play a key role here, establishing channels of communication, reconciling countries' differing approaches, and strengthening transparency and cooperation. As with the World Trade Organization, the International Court of Justice, and the International Criminal Court, their mission is to ensure compliance with established rules and to resolve disputes that may arise in fields where national interests often compete. Some, therefore, advocate for the creation of a new ad hoc authority. In that event, what institutional innovations would be necessary to better coordinate the regulation of AI?

Several precedents have been put forward as potential models. Firstly, AI, similarly to nuclear technology, can be defined as a "versatile" or "general-purpose" technology, due to three key characteristics: its high potential for continuous improvement, its pervasiveness in many industrial sectors, and its complementarity with other technologies. Some lessons can therefore be learned from nuclear regulatory bodies.

155. "OECD and UN announce next steps in collaboration on Artificial Intelligence", Press release, OECD, September 22, 2024, available at: www.oecd.org.

Firstly, where research is concerned, several associations, along with the Swiss Federal Institute of Technology in Zurich, are campaigning for the creation of a "CERN for AI". Located a short distance from Geneva, the European Organization for Nuclear Research (CERN) is the largest particle physics center in the world. It is an open-source, international public research infrastructure. Whereas CERN has particle accelerators, the research center proposed by the Large-Scale Artificial Intelligence Open Network (Laion) would be equipped with machines containing some 100,000 accelerators[156] (e.g., graphics processing units – GPUs). These would be run by participating states and could be used by international researchers, in accordance with authorization requirements, similar to biological research laboratories. All results would then become public.

Such an infrastructure would have the advantage of freeing AI research from the hold of large multinationals. It would not, however, address the need for oversight and verification mechanisms, essential to a regulatory body's proper functioning. Returning to the example of nuclear policy, some have suggested a model similar to the International Atomic Energy Agency (IAEA).

The IAEA, based in Vienna, was established in 1957 under the auspices of the UN and promotes the safe, secure, and peaceful use of nuclear technologies, while monitoring possible violations of the 1968 Treaty on the Non-Proliferation of Nuclear Weapons (NPT). With its 178 member states, its worldwide presence, and the support of the major nuclear powers, it works to develop the use of nuclear energy for electricity production and limit its military applications.

AI being a dual-use technology, such an approach could prove worthwhile. Indeed, combating the very real threats associated with AI technologies should not serve as a pretext to suspend the many benign and even beneficial uses of these technologies.[157] An agency capable of addressing both sides of the issue would therefore be a major asset, particularly as discussions on the international governance of "civilian" and "military" AI are currently conducted entirely separately, which makes little sense given the potential for "civilian" applications to be misused for malicious purposes.

This idea is in fact proving very popular, with politicians like Rishi Sunak, tech executives like Sam Altman, but also with Antonio Guterres. Yet it certainly would not be easy to implement. In the case of the IAEA,

156. R. Karayan, "IA générative : une association allemande milite pour une recherche ouverte sur le modèle du CERN", *L'Usine digitale*, April 6, 2023.
157. The same is true for chemical weapons. This is why the mission of the Organization for the Prohibition of Chemical Weapons is not only to destroy existing arsenals and related military installations throughout the world, but also to monitor specific activities in the chemical industry to ensure that there is no risk of military misappropriation, and to promote cooperation between States for the peaceful use of chemistry.

which was established in 1957, it was not until the NPT came into force in 1970 that the agency was able to effectively monitor the nuclear weapons programs of participating countries and enforce safety standards. Today, talks on AI governance do not sufficiently recognize the essential role of treaties. In their eagerness to create new international institutions, political leaders tend to forget that actual enforcement can only proceed from binding agreements by states, and can only be guaranteed when accompanied by mechanisms imposing sanctions when violations occur.

In a book on the IAEA, historian Elisabeth Roehrlich highlights two key factors that have made the IAEA's nuclear safeguards effective: the legal agreements between the agency and its member states and the technical tools used to ensure compliance with these agreements.[158] Likewise, robust AI governance must involve both new regulations and the resources and technical capabilities to enforce them.

International regulatory bodies can only carry out their mission successfully if their mandate and the established regulations are clear and enforceable, both for businesses and governments. Policymakers must therefore begin by outlining these laws' prerequisites and their content before creating the agencies responsible for enforcing them. AI, with its speed of development, its opacity, and its constant evolution, is fundamentally different from other technologies, creating the need for a new approach to international oversight. Rather than being intimidated by this challenge, regulators should see it as an opportunity for innovation and creativity in the design of an effective regulatory framework.

Another interesting example is the Intergovernmental Panel on Climate Change (IPCC), an organization that is often cited as a potential model for AI. Rishi Sunak, along with other personalities like former Google CEO Eric Schmidt, say they have drawn inspiration from the IPCC, which compiles scientific research on climate change and organizes the annual Conference of the Parties (COP). Even before the United Kingdom organized the first summit on AI security, plans for an "IPCC of AI" had a clear mandate, however: the organization would not issue policy recommendations. Its role would rather consist of regularly compiling AI research, identifying common concerns, and presenting policy options, without giving any explicit directives. This limited approach, therefore, does not appear to offer a path to a binding treaty, with strong guarantees and real constraints on the influence of tech companies. An "IPCC of AI", as an independent research group, would certainly have its uses, especially given the opacity of the information firms are providing on the technology, and the importance of basing regulations on scientific consensus. However,

158. E. Roehrlich, *Inspectors for Peace: A History of the International Atomic Energy Agency*, Baltimore: Johns Hopkins University Press, 2022.

facilitating research is only a preliminary factor for the development of regulations, without which effective AI governance cannot be achieved.

In drawing inspiration from organizations such as the IAEA or the IPCC for AI governance, we risk overlooking the unique and novel challenges posed by this technology. Unlike nuclear technologies, which are mainly under government control, AI capabilities are concentrated in the hands of a few companies which actively market their products. They are also much less costly and difficult to produce than it is, for example, to enrich uranium in order to develop a nuclear weapon. AI technologies can therefore be disseminated extremely widely, including to non-state and malicious actors.

Translating key principles into technical terms

It is not enough to create a regulatory body to uphold the principles being established with regard to AI. These key principles must also then be made applicable in practice. While they may provide guidance for the development of trustworthy AI, and offer policymakers recommendations that will help them to formulate effective policies, they remain general in nature and leave a great deal open to interpretation. For the resulting safeguards to be robust, standards must first be harmonized, and the principles on which the various stakeholders have agreed must then be translated into technical terms. International standardization bodies will therefore be instrumental in bringing about AI governance that goes beyond wishful thinking and is capable of effectively holding the American and Chinese digital giants to account.

Technical standards are essential for defining the parameters of AI systems, whether these concern basic reference architectures, security and ethical requirements, or the technical functionalities of specific applications in a wide range of fields, e.g., healthcare, education, advanced manufacturing, energy and agriculture. Both China and the EU have sought to establish technical standards to mitigate risks and achieve general policy objectives in their efforts to control and guide the development of AI.[159]

This process of conversion results in fierce competition, however, in what Paul Timmers has called the "geopolitics of standardization",[160] that is, the development of technical standards governing the use of computer systems. This competition takes place at the national, regional – notably European – and international level. The governance of standardization

159. C. Wang and Y. Yin, "China Launches Global AI Governance Initiative, Offering an Open Approach in Contrast to US Blockade", *Global Times*, October 18, 2023.
160. P. Timmers, "Géopolitique de la normalisation", *Le Grand Continent*, June 2, 2020, available at: https://legrandcontinent.eu.

organizations is rather unusual, however. To get a seat at the European Telecommunications Standards Institute (ETSI), for example, "you just have to pay, and the higher the amount, the more votes you get".[161] As a result, while the major Chinese and American digital companies are over-represented, European companies are rare, given the high cost of entry.

For their part, the European Committee for Standardization and the European Committee for Electrotechnical Standardization (CEN-CENELEC) adopt standards set by the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC), which each have one representative per state, but whose seats are, here too, claimed by the digital giants. Thus, the representatives most active in AI standardization are currently employees at Microsoft, IBM, Google and Huawei, since representatives are appointed based on their nationality, regardless of the company they work for. In 2022, the Irish and Austrian heads of delegation to the ISO committee on AI standardization were Huawei employees, while the British and German heads of delegation were employed by Microsoft. In a job description presented by Huawei to one of these experts, the objectives listed included: "countering European legislation".[162]

The result is that European and global standards for AI systems are primarily being elaborated by private corporations, with clear biases. Beyond the obvious issue of sovereignty this raises, these private actors may also be acting in a manner not incompatible with some state strategies. As Paul Timmers has written about China:

> "This is a deliberate strategy by the Chinese government to write the rules of the game in emerging fields of information and communication technology (ICT) and to break with past standards, which were largely defined by the United States and Europe."[163]

During the implementation of the AI Act, for example, the European Commission issued a number of directives whose implementing decrees were ultimately defined by standardization work. When the Commission states, for example, that AI must be "robust", it is the technical translation of this adjective in international standardization forums that establishes the rule. "Robustness" then becomes a standard applied to AI systems. Those responsible for drafting technical standards therefore wield extraordinary power, as certain countries, like the United States and China, have clearly understood. As Thierry Breton remarked in February 2022: "Technical standards hold strategic importance. Europe's technological sovereignty, its ability to cut its dependencies and to protect the values of the EU, will depend on our ability to set standards globally".

161. Author's research interview with a standardization expert.
162. Ibid.
163. P. Timmers, "Géopolitique de la normalisation", op. cit.

Additionally, beyond the competition over technical standardization in order to project power and values, it is unclear where these principles should be put into practice: in procedures or in objectives.[164] The prevailing approach has been to focus on results – the production of standards – by taking a potential problem caused by AI, and identifying governance principles likely to mitigate risks or increase the likelihood of reaching a desired outcome.[165] The alternative would be to focus on procedures by paying more attention to the way in which governance processes are carried out.

The principle of justice can therefore be understood as a procedural value (ensuring the fairness and representativeness of governance bodies) as well as a distributional outcome (ensuring that, technically, AI systems do not fuel injustices). For example, it is likely that AI-driven automation will cause certain populations to lose their jobs. These populations would then bear a disproportionate share of the technology's negative externalities, which would not be offset by access to its benefits.[166] To focus solely on justice as a procedural value, therefore, neglects the distributional impact of the spread of AI systems.

At a time when the key principles governing the development and deployment of AI are being established – trust, robustness, explainability, accountability, safety, security, inclusivity, sustainability, respect for democratic values and human rights, etc.[167] – the technical, procedural, and normative ramifications of these principles must be carefully considered.

Essential coordination with national regulations

Given that private actors within the AI ecosystem can operate in several national jurisdictions, regulatory efforts must also be cross-border. Only by introducing common rules will states be able to ensure that these companies are exposed to similar regulatory environments. Only this kind of approach can promote global development of AI and discourage companies from turning to countries with laxer regulations.

International governance will only be effective, however, if it is coordinated with national regulations. Promoting institutions to establish and enforce norms and standards without simultaneously establishing national regulations is at best naive, and at worst a deliberate calculation –

164. J. Tallberg et al., "The Global Governance of Artificial Intelligence: Next Steps for Empirical and Normative Research", op. cit.
165. A. Dafoe, *AI Governance: A Research Agenda*, op. cit.
166. E. Erman and M. Furendal, "Artificial Intelligence and the Political Legitimacy of Global Governance", op. cit.
167. The OECD AI principles are inclusive growth, sustainable development and well-being; respect for the rule of law, human rights and democratic values (including fairness and privacy); transparency and explainability; robustness, security and safety; and accountability.

an act of "false compliance" aimed at holding back competitors. Transposing international standards into national policies is essential to ensure the interoperability of standards between different jurisdictions and prevent further regulatory chaos. This is why the AI Action Summit aspires to build a robust governance framework based on the "convergence of AI-specific standards and public policies".

National regulatory efforts will, however, continue to attract the attention of the digital giants, requiring renewed vigilance. Eric Schmidt, for example, has invested considerable sums in AI start-ups and research companies, while advising the U.S. government on policy in this area, with unsurprising emphasis on corporate autonomy. This clear potential for conflicts of interest makes it necessary to establish legally enforceable safeguards prioritizing the public interest, rather than weak standards serving the financial interests of Big Tech.[168]

One final concern is the potential for AI applications to violate pre-existing laws on a wide range of subjects. For this reason, the supervisory bodies currently being established must have the power to enforce existing antitrust, non-discrimination, and intellectual property laws, which again makes coordination with national regulations a necessity.

Towards "future proof" governance

Beyond the issue of adopting and applying new standards to regulate AI, there is also the question of their potential obsolescence, given the dramatic technological advances that are sometimes made in this field. Standard-setting has traditionally followed one of four approaches.[169]

Firstly, existing rules can be reinterpreted to apply to AI.[170] For example, many are calling for the international humanitarian law principles of distinction, proportionality and precaution to be extended, through their reinterpretation, to apply to lethal autonomous weapon systems, without any alterations to the original legal text.

New AI regulations may also take the form of amendments to existing rules. In the case of autonomous vehicles, for example, the 1968 Vienna Convention on Road Traffic was amended in 2015 to include provisions relating to AI.[171]

168. M. Schaake, "The Premature Quest for International AI Cooperation", op. cit.
169. J. Tallberg et al., "The Global Governance of Artificial Intelligence: Next Steps for Empirical and Normative Research", op. cit.
170. M. Maas, *Artificial Intelligence Governance under Change: Foundations, Facets, Frameworks*, Copenhagen: University of Copenhagen, 2021.
171. M. Kunz and S. Ó hÉigeartaigh, "Artificial Intelligence and Robotization" in R. Geiss and N. Melzer (eds.), *The Oxford Handbook on the International Law of Global Security*, Oxford: Oxford University Press, 2020.

Regulation can also come in the form of an entirely new framework, either as a result of a new state practice transforming into customary international law, or through a regulatory innovation, such as a new legal act or a new treaty[172] – as is the case, for example, with the AI Act.

These different standard-setting methods do not, however, prevent innovations from slipping through legal loopholes, given the speed, scale and uncertainty associated with the development of AI. States also often make use of new technologies long before specific rules to regulate their use are agreed upon.[173] During this interval, they will develop uses for these technologies, and form ideas about what constitutes "appropriate" use, which will alter existing standards.

AI governance is therefore at risk of constantly lagging behind technological advances. No one can predict what AI will be capable of in the future, which is why the policies and institutions governing it must be designed to be flexible and adaptive, to stand the test of time and innovation. To this end, Article 97 of the AI Act allows the Commission to adopt delegated acts to update the regulations in order to account for technological developments. Similarly, the Council of Europe Framework Convention on AI aims to ensure that any legal vacuums that might result from overly rapid technological advances are avoided. Thus, in order to stand the test of time, the Framework Convention does not regulate technology and is essentially technologically neutral.

172. M. Maas, *Artificial Intelligence Governance under Change: Foundations, Facets, Frameworks*, op. cit.
173. R. Alcala and E. Talbot Jansen, *The Impact of Emerging Technologies on the Law of Armed Conflict*, New York: Oxford University Press, 2019.

Conclusion

At a time when AI is rapidly becoming an essential technology, many fundamental questions remain unanswered. How can we effectively protect users' privacy in the face of massive data collection? What mechanisms could ensure that algorithms involved in crucial decisions, such as access to employment or public services, are transparent and fair? And above all, how can we prevent these systems from straying down the wrong path, making it possible for AI to generate dangerous content or arm autonomous systems? These challenges highlight the urgent need to define common guidelines to ensure that this technology is used in an ethical and responsible manner.

Without coherent global governance, a technological arms race could cause us to repeat past mistakes. As happened during the nuclear arms race in the 20th century, states and private companies risk prioritizing economic and geopolitical rivalries to the detriment of our collective security.[174] To prevent such a scenario, policymakers must not only disentangle the potential benefits from the risks associated with AI, but also encourage development that maximizes the former while mitigating the latter.

The Paris summit in February 2025 and initiatives by institutions such as the United Nations, the G7 and the OECD will all be important steps in clarifying the global governance of AI. But a truly coordinated, risk-based approach, ensuring interoperability between the various regulatory frameworks, remains to be developed. Such an approach will be essential to prevent abuse, reduce inequalities between the different regions of the world, and ensure that the benefits of AI are shared equitably.

As the international community develops a better understanding of these new technologies, the priority must be given to concrete action rather than symbolic gestures. While summits, codes of conduct, regulations and declarations have drawn attention to the importance of AI governance, binding commitments will be needed to bring about real change. If we are to move beyond vague hopes and declarations of intent, we must develop mechanisms to verify compliance and impose sanctions should the rules be broken. In this respect, the AI Action Summit in Paris could provide momentum and enable the development of an "AI Paris Agreement".

It cannot be denied that multilateralism is losing ground at the moment, all the more so now that the United States under Donald Trump is withdrawing from one international agreement after another (World Health

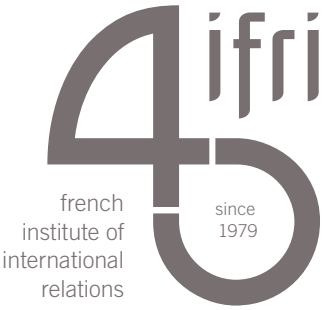
174. A. Dafoe, *AI Governance: A Research Agenda*, op. cit.

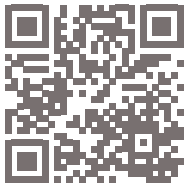
Organization, Paris Climate Agreement…). Rising to the challenge of regulating AI, as universal a problem as could be, must nevertheless be seen as an opportunity to breathe new life into our ailing multilateralism. The Council of Europe Framework Convention on AI is a promising step forward in this regard. Unlike other initiatives, this text is legally binding and offers a model to structure AI governance on a global scale. It is also an important milestone, as it is one of the first times that the United States and the EU have officially aligned themselves with respect to AI regulation.[175]

At the Summit of the Future in September 2024, the leaders of the 193 United Nations Member States also unanimously adopted the "Pact for the Future", which aims to reinvent the multilateral system, as well as the "Global Digital Compact", which should help them address long-term challenges in this area. Two new international mechanisms have also been created: an independent international scientific panel on AI and a global dialogue on AI governance. This could be the first building block in the complex regulatory framework needed to govern this technology.

To implement this ambitious agenda, an international body should be established to harmonize the various initiatives and assign responsibilities (ethical, security, societal, scientific, technical, commercial, etc.) to avoid redundancies and contradictions. The goal is not to start from scratch, but to coordinate what is already in place. Without such leadership, there can be no real global governance of AI.

175. R. Csernatoni, "The AI Governance Arms Race: From Summit Pageantry to Progress?", op. cit.





27 rue de la Procession 75740 Paris cedex 15 – France

Ifri.org