cerre Centre on Regulation in Europe

GLOBAL GOVERNANCE OF DIGITAL ECONOMY: ARTIFICIAL INTELLIGENCE

September 2024

Adrien Abecassis

GLOBAL GOVERNANCE FOR THE DIGITAL ECOSYSTEMS: PHASE TWO

cerre

Issue Paper

Global Governance of Digital Economy: Artificial Intelligence

Adrien Abecassis

September 2024

As provided for in CERRE's bylaws and procedural rules from its "Transparency & Independence Policy", all CERRE research projects and reports are completed in accordance with the strictest academic independence.

The project, within the framework of which this report has been prepared, received the support and/or input of CERRE member organisations. However, these bear no responsibility for the contents of this report. The views expressed in this CERRE report are attributable only to the authors in a personal capacity and not to any institution with which they are associated. In addition, they do not necessarily correspond either to those of CERRE, or of any sponsor or of members of CERRE.

info@cerre.eu – www.cerre.eu

TABLE OF CONTENTS

About CERRE

Providing top quality studies and dissemination activities, the Centre on Regulation in Europe (CERRE) promotes robust and consistent regulation in Europe's network and digital industries. CERRE's members are regulatory authorities and operators in those industries as well as universities.

CERRE's added value is based on:

- its original, multidisciplinary and cross-sector approach;
- the widely acknowledged academic credentials and policy experience of its team and associated staff members;
- its scientific independence and impartiality;
- the direct relevance and timeliness of its contributions to the policy and regulatory development process applicable to network industries and the markets for their services.

CERRE's activities include contributions to the development of norms, standards and policy recommendations related to the regulation of service providers, to the specification of market rules and to improvements in the management of infrastructure in a changing political, economic, technological and social environment. CERRE's work also aims at clarifying the respective roles of market operators, governments and regulatory authorities, as well as at strengthening the expertise of the latter, since in many Member States, regulators are part of a relatively recent profession.

About the Author

Adrien Abecassis is a seasoned diplomat and former advisor to the President of France, where he played a key role in shaping French policies on European affairs and providing strategic political advice. He currently serves as the Director of Policy at the Paris Peace Forum, an international platform dedicated to fostering multilateral cooperation and improving global governance. His academic credentials include fellowships at Harvard University and the University of California, Los Angeles, where he contributed to research in international affairs. Additionally, he was a fellow at the Belfer Center for Science and International Affairs at Harvard Kennedy School during 2020-21.

1. Introduction: Context of AI Governance

Artificial Intelligence (AI) has transitioned from a conceptual framework to a transformative force. Let's start with **5 facts**:

1. **We are on a clear trajectory towards creating non-human intelligence. It will eventually surpass human intelligence across a broad range of tasks. This is not a matter of if, but of when, and the timeline is measured in years, not in decades or generations. Each of these statements represents a profound shift compared to only 4-5 years ago.**

   This perspective is driven by AI's exponential learning curves and the continuous integration of vast datasets that fuel machine learning models.[1] Artificial General Intelligence (AGI), where machines possess the ability to understand, learn, and apply knowledge across a wide range of tasks, is becoming a focal point of research.[2] AGI would represent a critical threshold, beyond which machines could outperform humans in virtually every cognitive task.

2. **AI has the potential to catalyse transformative and unprecedented advancements – in science, engineering, health, education, and many other fields.**

   AI can enable breakthroughs in fundamental and applied research – discovering new drugs, materials, or physical phenomena, or solving complex and intractable problems, such as climate change or disease.[3] AI can also enhance our own capabilities and augment our intelligence, allowing us to achieve more than we ever imagined.[4] In engineering, AI can optimise design processes and improve manufacturing efficiency. In healthcare, AI can enhance diagnostics, personalise treatments, and streamline administrative tasks. In education, AI can provide personalised learning experiences and assist teachers in managing classrooms.

   These advancements could usher in a new Renaissance, similar to the period of significant human progress seen in the 14th to 17th centuries. Just as the original Renaissance was characterised by advancements in art, science, and culture, a new AI-driven Renaissance could lead to breakthroughs across a wide range of fields, transforming society in profound ways.

3. **However, this potential will not be realised spontaneously. Deliberate and strategic, well-crafted policies and institutions are essential.**

[1] Jaime Sevilla et al., "Compute Trends Across Three Eras of Machine Learning," in 2022 International Joint Conference on Neural Networks (IJCNN), 2022, 1–8, https://doi.org/10.1109/IJCNN55064.2022.9891914.

[2] Katja Grace et al., "Thousands of AI Authors on the Future of AI" (arXiv, April 30, 2024), https://doi.org/10.48550/arXiv.2401.02843.

[3] OECD, Artificial Intelligence in Science: Challenges, Opportunities and the Future of Research (Paris: Organisation for Economic Co-operation and Development, 2023), https://www.oecd-ilibrary.org/science-and-technology/artificial-intelligence-in-science_a8d820bd-en.

[4] Maithra Raghu and Eric Schmidt, "A Survey of Deep Learning for Scientific Discovery" (arXiv, March 26, 2020), https://doi.org/10.48550/arXiv.2003.11755.

Achieving an AI-driven Renaissance requires fostering and sustaining innovation. As Romer pointed out in his Nobel lecture,[5] innovation, as an engine of economic growth and social progress, is not inevitable. Innovation depends on the creation and diffusion of ideas. New ideas are non-rivalrous – one person's use does not diminish their availability to others – and partially excludable goods – access can be restricted but not entirely. These characteristics involve positive and negative externalities which public policies and institutions should balance through specific measures including favourable conditions as well as the necessary infrastructure and finance to enable innovation, ensuring widespread access and maximising societal benefits.[6] Trade-offs arise in the realm of protecting and sharing intellectual property (IP). This reverberates across the open-source / proprietary debate surrounding AI.

Without appropriate governance and regulation, AI risks failing to generate new advancements and could unintentionally lead to scenarios which undermine societal well-being and economic stability: from privacy breaches and unfairness in AI systems, to cybersecurity threats, job displacement, or market monopolies.[7] Effective policies are crucial to navigate these challenges.

4. **The advancements and benefits of AI are not evenly distributed and may create new forms of inequality and injustice within and between countries.**

AI could be a "great divider" if not well managed, exacerbating the existing gaps and disparities in access, opportunity, and power.[8] Within countries, AI could affect the structure and dynamics of the labour market, creating more demand for high-skilled workers and less for low-skilled ones.[9] This could lead to labour market polarisation and social exclusion, unless adequate measures are taken to support the reskilling and redeployment of workers, as well as the provision of social protection and safety nets.[10]

Between countries, AI could widen the gap between the Global North and the Global South, as the former may have more access and control over the essential technologies and resources for AI, while the latter may face more barriers and vulnerabilities.[11] This could increase the tensions and conflicts between different regions and groups, unless a more inclusive and cooperative approach is adopted.[12]

---

[5] Paul Romer, "Nobel Lecture: On the Possibility of Progress," February 5, 2019, https://paulromer.net/prize/.

[6] Erik Brynjolfsson and Gabriel Unger, "The Macroeconomics of Artificial Intelligence," IMF, December 2023, https://www.imf.org/en/Publications/fandd/issues/2023/12/Macroeconomics-of-artificial-intelligence-Brynjolfsson-Unger.

[7] Daron Acemoglu, "Harms of AI," in The Oxford Handbook of AI Governance, ed. Justin B. Bullock et al. (Oxford University Press, 2024), 0, https://doi.org/10.1093/oxfordhb/9780197579329.013.65.

[8] Kristalina Georgieva, "AI Will Transform the Global Economy. Let's Make Sure It Benefits Humanity.," IMF, January 14, 2024, https://www.imf.org/en/Blogs/Articles/2024/01/14/ai-will-transform-the-global-economy-lets-make-sure-it-benefits-humanity.

[9] David Autor, "Polanyi's Paradox and the Shape of Employment Growth" (Cambridge, MA: National Bureau of Economic Research, September 2014), https://doi.org/10.3386/w20485.

[10] Carlo Pizzinelli Li Augustus J. Panton,Marina Mendes Tavares,Mauro Cazzaniga,Longji, "Labor Market Exposure to AI: Cross-Country Differences and Distributional Implications," IMF, accessed July 8, 2024, https://www.imf.org/en/Publications/WP/Issues/2023/10/04/Labor-Market-Exposure-to-AI-Cross-country-Differences-and-Distributional-Implications-539656.

[11] Somya Joshi and Björn Nykvist, "Anticipating Futures: How Artificial Intelligence Acts as an Amplifier of Inequity," Https://Sdgs.Un.Org/, 2023.

[12] "Automation and AI: Implications for African Development Prospects?," Center For Global Development, accessed July 8, 2024, https://www.cgdev.org/publication/automation-and-ai-implications-african-development-prospects.

5. **This path is heavy in unpredictability and uncertainty. These are due to three major factors:**

Technological uncertainty: AI may involve non-linear developments and emergent properties[13] that are difficult to anticipate, leading to unexpected outcomes. We still do not fully understand the behaviours of AI systems.[14] In some scenarios, AI may evolve and adapt to changing environments and goals, challenging our ability to monitor and regulate it.

Geopolitical dynamics: Geopolitical events, changes in government policies, and international relations, can significantly impact AI development.[15] Tensions between major AI players like the US and China influence the direction and pace of AI advancements.[16] The occurrence of extreme events, such as political crises or military conflicts, may affect more dramatically the course and consequences of AI.

Societal reactions: Public perceptions and societal reactions to AI will also shape its development and deployment.[17] Fears of AI misuse for instance, such as concerns about job displacement and privacy violations, could lead to regulatory constraints and hinder progress. This could be all the more unpredictable, as a result, the trajectory and impact of AI on specific sectors (jobs, for instance) requires strategic foresight and scenario-planning without relying solely on past data and experiences: patterns and trends from previous iterations of AI systems may not be applicable or relevant to future generations, as the technology may undergo qualitative and quantitative changes.[18] The actions and reactions of various actors – governments, corporations, and individuals – will influence the direction and pace of AI development and deployment.

[13] Jason Wei et al., "Emergent Abilities of Large Language Models," Transactions on Machine Learning Research, June 26, 2022, https://openreview.net/forum?id=yzkSU5zdwD.
[14] Rylan Schaeffer, Brando Miranda, and Sanmi Koyejo, "Are Emergent Abilities of Large Language Models a Mirage?," 2023, https://openreview.net/forum?id=ITw9edRDlD.
[15] Eric Schmidt, "Innovation Power: Why Technology Will Define the Future of Geopolitics," Foreign Affairs 102 (2023): 38.
16 Amelia C. Arsenault and Sarah E. Kreps, "AI and International Politics," in The Oxford Handbook of AI Governance, ed. Justin B. Bullock et al. (Oxford University Press, 2024), 0, https://doi.org/10.1093/oxfordhb/9780197579329.013.49.
[17] Baobao Zhang, "Public Opinion toward Artificial Intelligence," in The Oxford Handbook of AI Governance, ed. Justin B. Bullock et al. (Oxford University Press, 2024), 0, https://doi.org/10.1093/oxfordhb/9780197579329.013.36.
[18] Yogesh K. Dwivedi et al., "Evolution of Artificial Intelligence Research in Technological Forecasting and Social Change: Research Topics, Trends, and Future Directions," Technological Forecasting and Social Change 192 (July 1, 2023): 122579, https://doi.org/10.1016/j.techfore.2023.122579.

2. Dimensions of AI: And Why It Prevents a Unified Regime

One of the difficulties in designing an AI governance framework is the linkage between three sets of factors that both individually and collectively influence the progress and outcomes of its implementation and enforcement. **AI is simultaneously:**

- **an economic competition**

  AI is a strategic asset that can confer competitive advantages to countries and firms that master it.

  It is an extremely capital-intensive competition, requiring very significant investment in research, development, and infrastructure. Companies (and governments) must make strategic decisions about how much to invest in AI. Within the industry, a shift occurs from operational expenditure to capital expenditure, as companies need to acquire more hardware and software assets. This can create high-cost barriers to entry and innovation and may lead to an increase in market power concentration among a handful of dominant players.

  In a context of monetary and fiscal tightening, the economic dynamics of AI become even more pronounced. Public actors have to make trade-offs and allocate scarce resources among different domains and sectors. For public actors, investment in AI is increasingly competing with other priorities, such as climate change and defence, which are also capital-intensive.

  The energy demands of AI, particularly for training large models, add another layer of complexity. Data centres and computational resources consume vast amounts of electricity (and water), adding to the substantial operational costs and environmental considerations. The electricity demand for AI computation is expected to reach at least 70 TWh in 2026, which is comparable to the consumption of smaller European countries.[19] As a result, having abundant and cheap energy becomes an even more important economic advantage, and an attractiveness factor for AI investments. Countries or regions that have low-cost and reliable energy sources can attract AI actors who seek to lower their operational costs. This dynamic will lead to the emergence of new players with immense capital and energy resources, such as those in the Gulf region. Consequently, this can reposition the supply chains and partially draw a new geography of the AI economy.

  This economic competition has the potential to drive Schumpeterian creative destruction, where old industries are replaced by new, more efficient ones – leading to economic growth and innovation. But the disruptive effects on the structure and dynamics of markets and industries could create social challenges that would in turn require investing in education and training to equip workers with the skills and competencies that are in demand in the AI economy. Facilitating the transition of workers into new or different jobs requires reskilling

[19] International Energy Agency, "Electricity 2024 - Analysis and Forecast to 2026," 2024.

and upskilling programs, as well as financial and social support for those who are displaced or affected by AI.[20]

The social impact of AI may be difficult to foresee – especially on the labour market – in terms of number, quality, and skills of jobs. Automation and AI-driven processes can cause job loss, especially in sectors like manufacturing and retail, and generate new possibilities in tech-driven fields. Different studies suggest that millions of jobs are at stake – but predictions differ greatly from one study to another.[21] Significant job automation is predicted to impact routine tasks, potentially hollowing out labour markets and polarising incomes,[22] with some experts cautioning that AI might displace more jobs than it creates unless policies are implemented to foster job creation.[23] Others argue that the impact of automation on job displacement is overstated, emphasising the complementarities between automation and human labour that enhance productivity, increase earnings, and boost labour demand.[24]

The challenge remains in overseeing this change and ensuring that workers are reskilled for new roles.

- **a geopolitical confrontation**

  AI is also a source of power and influence that can alter the balance and relations between states and regions, with AI at the epicentre of a technological race and a geopolitical rivalry between the US and China.

  The two leading powers have different visions and values for the development and use of AI and are also competing for global leadership and influence in setting the norms and standards for AI governance. **The race for AI supremacy is intensifying an East/West fragmentation**. Both nations are investing heavily in AI research and development. This competition extends beyond economics, impacting national security and global influence, as AI capabilities are seen as essential for maintaining technological and military superiority.

  Furthermore, **AI may also exacerbate the North/South polarisation**, with developed countries advancing rapidly in AI while the Global South lagging behind.[25] This disparity can lead to a widening technological gap, exacerbating global inequalities. The lack of access and participation in AI development and deployment may limit the opportunities and benefits for the Global South and expose them to the risks of AI.

- **an ethical debate**

[20] Andrew Berg, Chris Papageorgiou, Maryam Vaziri, "Technology's Bifurcated Bite," IMF, December 2023, https://www.imf.org/en/Publications/fandd/issues/2023/12/Technology-bifurcated-bite-Berg-Papageorgiou-Vaziri.
[21] Mauro Cazzaniga Tavares Florence Jaumotte,Longji Li,Giovanni Melina,Augustus J. Panton,Carlo Pizzinelli,Emma J. Rockall,Marina Mendes, "Gen-AI: Artificial Intelligence and the Future of Work," IMF, accessed July 8, 2024, https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2024/01/14/Gen-AI-Artificial-Intelligence-and-the-Future-of-Work-542379.
[22] Carl Benedikt Frey and Michael A. Osborne, "The Future of Employment: How Susceptible Are Jobs to Computerisation?," Technological Forecasting and Social Change 114 (January 1, 2017): 254–80, https://doi.org/10.1016/j.techfore.2016.08.019.
[23] Daron Acemoglu and Simon Johnson, "Rebalancing AI - Daron Acemoglu Simon Johnson," IMF, December 2023, https://www.imf.org/en/Publications/fandd/issues/2023/12/Rebalancing-AI-Acemoglu-Johnson.
[24] David H. Autor, "Why Are There Still So Many Jobs? The History and Future of Workplace Automation," Journal of Economic Perspectives 29, no. 3 (August 1, 2015): 3–30, https://doi.org/10.1257/jep.29.3.3.
[25] Nur Ahmed and Muntasir Wahed, "The De-Democratization of AI: Deep Learning and the Compute Divide in Artificial Intelligence Research" (arXiv, October 22, 2020), https://doi.org/10.48550/arXiv.2010.15581.

AI is not only a technical or economic issue, but also a moral and political one that raises fundamental questions about the values and principles.

Effective AI governance requires navigating complex policy trade-offs. One of them is how to balance efficiency and accuracy in decision-making with autonomy and human control. AI systems can process large amounts of data and generate predictions or recommendations faster and more accurately than humans, which can enhance productivity, quality, and innovation. But the same AI systems may also pose challenges to human autonomy, agency, and dignity, if they replace or override human decisions without sufficient transparency, explanation, or consent.

Such trade-offs arise across policy domains. AI could enhance public order and efficiency in policing or justice, helping detect and prevent crimes, optimising resource allocation, expediting legal processes, and reducing human errors and biases. However, these applications of AI may also raise serious concerns about the impact of surveillance on human rights and freedoms, enabling mass and indiscriminate data collection and analysis, facilitating facial recognition and biometric tracking, amplifying social profiling and discrimination. Every country is faced with the challenge of harnessing the positive potential of AI while minimising the risks and ensuring respects of their collective preferences.

One of the most debated aspects of AI is the possibility that, while surpassing human intelligence and control, it may create existential risks for humanity and the planet. Although the possibility of existential risks is contested, concerns should be earnestly deliberated. They should not be dismissed as irrational or sensationalist, nor should they be embraced uncritically or dogmatically. Concerns should be taken seriously, with genuine consideration for each argument, weighing the evidence and uncertainties, and acknowledging the ethical and social implications. This conversation already has implications for how we design, develop, and deploy AI systems, and how we govern their use and impact. In any case, it challenges us to think about what kind of future we want, and what kind of values and principles we want to uphold and promote.

In terms of governance architecture, this implies a clear outcome: we should forgo convergence and organise coexistence.

- **Economic rivalry is too intense to foster widespread cooperation**. Countries which fear being left behind or exploited by others have an incentive to defect from cooperation and pursue their own AI interests, even if this results in a suboptimal outcome for all. On paper, the US and China would benefit from sharing standards and best practices to improve the quality and safety of AI systems, but they also fear that doing so, beyond a baseline of core safety, would undermine their relative position and expose them to risks of espionage or sabotage. Economic rivalry bars convergence and challenges coexistence.
- **National security concerns further hinder convergence**. As AI systems pose new challenges and opportunities for military and intelligence activities, countries tend not to trust each other to use them responsibly or transparently. Some of these concerns may be exaggerated or used as a pretext to justify protectionist or isolationist policies, but they are a political fact and cannot be dismissed.

- **Cultural diversity is another obstacle to convergence**. AI systems reflect the values and preferences of their creators and users, and these may vary significantly across countries and regions. There is no universal agreement on what constitutes ethical, fair, or human-centric AI, and different cultures may have different expectations and concerns about the role and impact of AI in society. Some countries may have stricter regulations on privacy and data protection, while others may have more permissive attitudes towards facial recognition or social scoring. Some countries may have higher or lower tolerance for errors or biases in AI systems, depending on their legal and cultural traditions. These collective preferences and values are legitimate and must be respected.

  Countries will not be able to enforce their own values upon others, even if they endeavour to. The EU will not make China adopt its surveillance standards, nor make the US align its free speech rules and the same is true for the inverse. Countries will retain different views on the acceptable use of AI for surveillance or cyberwarfare and may not agree on common rules or oversight mechanisms.

  Moreover, AI systems may affect cultural diversity itself, by influencing language, identity, communication, and social norms. Therefore, global governance of AI should respect and protect cultural diversity and avoid imposing one-size-fits-all solutions.

- **If the concept of a single, global, regime for AI will remain illusory, we suggest that a more realistic approach to global governance of AI is not to seek convergence, but to organise coexistence.**

  This means acknowledging and respecting the diversity of economic, security, and ethical preferences that shape the development and use of AI systems and finding ways to accommodate them without creating unnecessary conflicts or interference. It also means identifying and addressing the spill overs and externalities of AI across borders (such as cyberattacks or environmental impacts) and ensuring that the benefits and risks of AI are equitably distributed among all stakeholders.

  Organising coexistence is not the ideal solution, but it may be the best approximation of the global optimum that can be achieved under the current circumstances.

A global optimum for AI governance would have three components:

- **Rapid advancement of technology**, given the possible immense and widespread benefits that AI can offer for human well-being, social progress, and economic growth. AI governance should foster innovation and experimentation and avoid unnecessary barriers or restrictions to the development and deployment of AI systems.
- **Safe advancement of technology**, given the potential risks that AI can pose for human rights, democracy, security, and stability. AI governance should ensure that AI systems are trustworthy, transparent, accountable, and aligned with human values and interests. It should also prevent and mitigate the negative impacts of AI on individuals, groups, and society, and protect against malicious or harmful uses of AI.
- **Fair access to technology**, given the impact that it would have if some countries or regions benefited from AI to the detriment of others, especially those in the Global South. AI governance should promote the inclusion and participation of all stakeholders and address

the inequalities and injustices that AI may create or exacerbate across the world. It should also facilitate the sharing and distribution of AI knowledge, data, and resources, and respect the sovereignty and autonomy of all countries and regions.

AI governance should strive to approach these criteria as best as possible, while recognising the trade-offs and tensions that may arise among them. It should also be adaptive to the evolving nature and challenges of AI.

3. Elements of AI Governance: Lacking a Globally Coherent Regime

There is no sign of a coherent and coordinated regime that can effectively address the various ethical, legal, and social implications of AI development and deployment. On the contrary, the current landscape of AI governance is characterised by fragmentation and diversity. A few key factors are shaping this landscape.

1. **Many aspects of AI governance are and will remain handled at the national or regional level, reflecting the specific needs and priorities of each jurisdiction.** Any architecture will have to build on this layer.

   Regulatory models adopted by the EU, China, the United States, and India illustrate their distinct political and cultural priorities. The EU's comprehensive, rights-driven regulations contrast with the US's flexible, innovation-centric approach. China's rigorous, state-controlled model underscores a focus on stability (and censorship) while India's framework aims to balance innovation with protective oversight, while being development-oriented.

Approaches of AI by the EU, China, the US, and India
All four regions/countries recognise the importance and potential of AI for economic and social development and have formulated national or regional AI strategies to guide their vision and actions. They all acknowledge the need to address the ethical, legal, and social challenges posed by AI, and have adopted principles or norms for the responsible development and use of AI. However, they differ in their emphasis and scope of AI policy and regulation – reflecting their values, interests, and capabilities.

Region	Regulatory Model	Features	Main texts
EU	Risk-based and rights-driven. Aims to foster human-centric, trustworthy, and ethical AI that respects fundamental rights, democracy, and the rule of law.	• Emphasises the protection of fundamental rights, democracy, the rule of law, and environmental sustainability. • Classifies AI applications into high-risk and low-risk categories, with stricter rules for the former. • Bans certain AI applications that are deemed unacceptable, such as social scoring and biometric categorisation. • Imposes fines and sanctions for non-compliance.	• The EU AI Act (AIA), adopted in March 2024, is the first comprehensive and binding legal framework for AI in the EU. • The General Data Protection Regulation (GDPR), adopted in 2016, sets high standards for data privacy and protection. • The Digital Services Act and the Digital Markets Act, adopted in 2022, aim to regulate online platforms and ensure fair competition.

China	State-driven and control-oriented. Assertive AI strategy to become a global leader and innovator in AI, and leverage AI for economic growth, social governance, and national security.	• Prioritises the development and deployment of AI for national security, social stability, and economic growth. • Grants the state full authority and access over AI data, algorithms, and applications. • Restricts the use and export of AI for sensitive or harmful purposes, such as subversion of state power or fake and harmful information. • Supports the AI industry through subsidies, infrastructure, and talent development.	• The New Generation Artificial Intelligence Development Plan, issued in 2017, outlines China's vision and strategy for becoming a global leader in AI by 2030. • The Interim Measures for the Management of Generative AI Services, enacted in 2023, regulate the development and use of generative AI services, such as chatbots and image generation. • The Data Security Law, passed in 2021, establishes a data classification system and imposes penalties for data breaches and misuse.
US	Market-driven and innovation-friendly. Flexible and market-driven approach to promote responsible innovation, foster public trust, and ensure national competitiveness and security in AI.	• Favours a limited government role and a self-regulatory approach by the private sector. • Supports the freedom of speech and technological innovation. • Relies on voluntary commitments and non-binding measures. • Adopts a sector-specific and context-based approach to AI regulation.	• The Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, issued in 2023, directs government agencies to prepare assessments and recommendations on AI in their domains. • The Voluntary Safeguards for Advanced AI Systems, agreed in 2023, set out best practices for AI safety, security, and ethics by seven leading AI companies. • The National Artificial Intelligence Initiative Act, enacted in 2020, authorises and coordinates federal investments and activities in AI research and development.
India	Development-oriented. Aims to leverage AI for economic development, social empowerment (especially for the underprivileged sections of society), and global cooperation.	• Pursues a balanced approach that encourages domestic innovation. • Promotes a multilateral and inclusive framework for AI governance that reflects the needs and values of the Global South. • Enhances AI literacy and capacity building among public and private stakeholders.	• The National Strategy for Artificial Intelligence, released in 2018, identifies five priority sectors for AI intervention: healthcare, agriculture, education, smart cities, and smart mobility. • The Digital India Act, proposed in 2023, covers various aspects of AI, data governance, and cybersecurity.

2. In addition to these regulatory frameworks, **national governance systems are aiming to encourage a set of essential factors** to create a favourable environment for AI development and deployment:

- **Expanding compute capacity:** National and regional policies around the globe are increasingly aimed at expanding compute capacity for AI, recognising its crucial, but extremely unevenly distributed, role in driving technological advancement and economic growth.[26] Many countries are building or upgrading domestic infrastructure – data centres, cloud services, and high-performance computing facilities – as well as shaping policies and incentives accordingly. In the United States, the CHIPS and Science Act allocates $52.7 billion for semiconductor R&D and manufacturing.[27] The EuroHPC Joint Undertaking focuses on developing the EU supercomputing services.[28] China has set up a 30% growth goal in national compute capacity by 2025.[29] The UAE's National AI Strategy 2031 is supported by an investment fund that could potentially reach $100 billion, along with partnerships involving entities like G42, a state-backed enterprise focused on developing AI infrastructure and applications.[30]
- **Energy:** Supporting this compute capacity necessitates significant increases in energy resources. According to recent estimates, data centres worldwide currently consume about 1-2% of global electricity, with projections suggesting this could rise to 3-4% by the end of the decade.[31] To meet this growing energy demand, countries and regions with abundant energy resources are positioning themselves as attractive locations for data centres. In particular, the Gulf states have been at the forefront of this trend, leveraging their energy abundance and strategic geographical location to attract major tech companies.
- **Investment:** AI innovation relies on the availability and allocation of financial resources, from both public and private sectors. Some regions – including the EU – face challenges to leverage sufficient investment (in the EU's case, a fragmented market, conservative risk-taking behaviours, and insufficient funding for start-ups). Many national governance schemes seek to foster a favourable investment climate for AI by providing incentives and reducing barriers.
- **Talent:** AI talent is scarce and in high demand, as it requires a mix of skills and expertise from various disciplines. There is a global competition to attract and retain skilled individuals, who can contribute to AI research and innovation. Investing in education and training programs, both formal and informal, is seen as critical by a growing number of countries, as well as

[26] OECD, "A Blueprint for Building National Compute Capacity for Artificial Intelligence," February 28, 2023, https://www.oecd.org/en/publications/a-blueprint-for-building-national-compute-capacity-for-artificial-intelligence_876367e3-en.html.
[27] Lennart Heim, Markus Anderljung and Haydn Belfield, "To Govern AI, We Must Govern Compute," 28 March 2024, n.d., https://www.lawfaremedia.org/article/to-govern-ai-we-must-govern-compute.
[28] "AI: Council Reaches Political Agreement on the Use of Super-Computing for AI Development," Consilium, May 23, 2024, https://www.consilium.europa.eu/en/press/press-releases/2024/05/23/ai-council-reaches-political-agreement-on-the-use-of-super-computing-for-ai-development/.
[29] "China Plans 30% Growth in National Compute Capacity by 2025," Tasnim News Agency, July 8, 2024, //www.tasnimnews.com/en/news/2024/07/08/3118923/china-plans-30-growth-in-national-compute-capacity-by-2025.
[30] Jayde Cheung, "Middle East Ramps up Bid to Become Global AI Hub," June 14, 2024, https://www.theasset.com/article/51696/middle-east-ramps-up-bid-to-become-global-ai-hub.
[31] "AI Is Poised to Drive 160% Increase in Data Center Power Demand," Goldman Sachs, June 28, 2024, https://www.goldmansachs.com/intelligence/pages/AI-poised-to-drive-160-increase-in-power-demand.html.

policies aimed at attracting the mobility and diversity of AI talent, by removing visa and immigration hurdles.

- **Data:** Data is an essential fuel for AI. However, not all data is equally valuable and accessible. Some data sets, such as health data, are highly coveted and sensitive, as they can provide benefits and insights for various domains. Most of them also pose risks and challenges for privacy and security.

3. **In each of these domains, one can expect bilateral and sometimes multilateral sectoral deals, similar to the global trade system.**

   Like bilateral free trade agreements, countries will eventually form bilateral or regional AI governance pacts tailored to their specific technological and ethical standards. These agreements would facilitate the exchange of AI technologies, data, and expertise, while addressing mutual concerns like data privacy, algorithmic transparency, and ethical AI use.

   These agreements would serve as intermediaries between bilateral and multilateral frameworks. They would converge AI policies within specific regions, fostering cooperation among geographically and culturally aligned nations. As seen in trade, Richard Baldwin's "domino effect" theory suggests that successful regional agreements could inspire broader participation and integration, eventually contributing to global AI governance.

   Bilateral agreements could act as testing grounds for innovative policies, which, if successful, could be scaled up to regional and eventually multilateral levels; and regional agreements would bridge the gap between local experimentation and global standardisation, ensuring that region-specific concerns are addressed within a broader framework.

   The sum of these pacts will form a web of arrangements, which could form the foundation of larger deals. But they will not naturally lead to a unified global governance architecture.

4. **It is already too late for a well-coordinated architecture or global bodies that would organise the coexistence of different AI systems and actors.**

   An ambitious proposal for a global AI governance mechanism is the Digital Stability Board (DSB), which was put forward in November 2022 by CERRE.[32] Drawing inspiration from the Financial Stability Board, which was established by the G20 after the 2008 global financial crisis and has been successful in setting global standards and norms, fostering policy coherence and consistency, and ensuring the resilience and stability of the international financial system, a DSB would have aimed to address the systemic and cross-border challenges posed by AI, facilitating coordination and information sharing.

   But today, more than two years after the CERRE proposal, the international landscape has become too crowded to make room for a coordinating body of this kind. A regime complex: i.e., a collection of non-hierarchical institutions that partially overlap and have various

[32] "Global Governance for the Digital Ecosystems", CERRE, https://cerre.eu/publications/global-governance-for-the-digital-ecosystems/.

functions and memberships, is therefore the most suitable governance model for AI.[33] This would allow for diversity and pluralism, as different values and norms can exist and compete, without forcing a uniform solution. The downside is that it can increase fragmentation and inconsistency, as different institutions may have goals and rules that conflict or coincide, creating gaps or clashes in the governance system.[34]

Multilateral Initiatives on AI Global Governance

Numerous multilateral initiatives have emerged as to establish guidance, principles, and standards for responsible and trustworthy AI. They are providing essential building blocks for an AI governance architecture, setting foundational elements that could support future frameworks. Nonetheless, most of these efforts face challenges in achieving widespread adoption and effective regulation due to their limited jurisdiction and voluntary nature.

Some of the most notable examples are:

- **The OECD Principles on Artificial Intelligence**,[35] adopted in 2019, which provide a set of recommendations for ensuring that AI is designed, developed, and used in a way that respects human values and dignity, inclusiveness, transparency, accountability, and security.
- **The UNESCO Recommendation on the Ethics of Artificial Intelligence**,[36] adopted in 2022, which establish the first global normative framework for ethical AI, covering issues such as human dignity, autonomy, justice, privacy, diversity, and sustainability.
- **The G7's Hiroshima Process**,[37] launched in 2022, which developed international guiding principles for human-centric AI, based on shared values and best practices. The 2024 **G7** reaffirmed its commitment to collaborate with the OECD on developing instruments to oversee the application of the Code of Conduct while expanding participation among key stakeholders in its **Digital Ministerial Declaration.**[38]
- **The International Telecommunication Union** (ITU)'s work on AI governance and standardisation,[39] including the annual AI for Good Global Summit, the development of numerous AI-related standards with a focus on preventing harmful behaviours as well as support for capacity building and policy assistance for developing countries.

[33] Emma Klein and Stewart Patrick, "Envisioning a Global Regime Complex to Govern Artificial Intelligence," Artificial Intelligence, Carnegie Endowment for International Peace, March 2024, https://carnegieendowment.org/research/2024/03/envisioning-a-global-regime-complex-to-govern-artificial-intelligence.
[34] Ian Bremmer and Mustafa Suleyman, "The AI Power Paradox," Foreign Affairs, August 16, 2023, https://www.foreignaffairs.com/world/artificial-intelligence-power-paradox.
[35] "Recommendation of the Council on Artificial Intelligence," OECD Legal Instruments, accessed July 18, 2024, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.
36 "Recommendation on the Ethics of Artificial Intelligence," UNESCO Library, accessed July 18, 2024, https://unesdoc.unesco.org/ark:/48223/pf0000381137.
37 "The Hiroshima AI Process: Leading the Global Challenge to Shape Inclusive Governance for Generative AI," The Government of Japan, accessed July 18, 2024, https://www.japan.go.jp/kizuna/2024/02/hiroshima_ai_process.html.
[38] "G7 Industry, Technology and Digital Ministerial Meeting" (G7 presidency, n.d.), https://assets.innovazione.gov.it/1710505409-final-version_declaration.pdf.
[39] "Moving AI Governance from Principles to Practice," ITU, April 19, 2024, https://www.itu.int/hub/2024/04/moving-ai-governance-from-principles-to-practice/.

- The United Nations has established a **High-Level Advisory Body on Artificial Intelligence** (HLAB on AI),[40] launched by the Secretary-General on October 26, 2023. This body comprises 38 experts tasked with making preliminary recommendations on AI governance, focusing on three main areas: international governance of AI, a shared understanding of risks and challenges, and identifying key opportunities and enablers. These recommendations will contribute to the preparations for the Summit of the Future and the negotiations on the **Global Digital Compact** (GDC).[41] The GDC negotiations are taking place in New York with the aim of adopting the compact in September 2024 during the Summit of the Future.

5. **Sectoral convergence on some key preferences will likely continue be sought between the like-minded, relying on a multi-stakeholder method.**[42]

   One of the emerging trends in AI governance is the development of sectoral or thematic initiatives that aim to align the preferences and standards of like-minded countries or regions on specific AI issues. These initiatives, covering a range of domains, could create convergence on key preferences for AI among major markets and stakeholders, and ease dilemmas for companies to deal with undefined or diverse regulations. However, they would not address the global and cross-cutting challenges of AI and would remain limited to a collection of (more or less) narrow issues.

Targeted Objectives
Numerous multi-stakeholder initiatives address significant global challenges. Below are selected examples highlighting efforts in various domains:
Fighting Extremist Content: The Christchurch Call,[43] a voluntary pledge by 50+ governments and a dozen major tech companies, aims to eliminate terrorist and violent extremist content online. Launched in 2019, after the attacks on mosques in Christchurch, New Zealand, which were livestreamed on social media, the pledge aims to prevent the abuse of digital technologies by terrorists and extremists while nonetheless safeguarding freedom of expression. The Christchurch Call has recently disclosed plans to evolve into a permanent Foundation. **Protecting Democracy**: The Partnership on Information and Democracy[44] is an intergovernmental agreement spearheaded by Reporters Without Borders and 38 countries, which commits to promote and protect the democratic principles and values in the online information and communication space. Adopted during the 2019 G7 Summit in Biarritz, France, it builds on the

[40] "AI Advisory Body," United Nations, accessed July 18, 2024, https://www.un.org/ai-advisory-body.
[41] "Global Digital Compact," Office of the Secretary-General's Envoy on Technology, accessed July 18, 2024, https://www.un.org/techenvoy/global-digital-compact.
[42] K. Gretchen Greene, "AI Governance Multi-Stakeholder Convening," in The Oxford Handbook of AI Governance, ed. Justin B. Bullock et al. (Oxford University Press, 2024), 0, https://doi.org/10.1093/oxfordhb/9780197579329.013.6.
[43] "The Christchurch Call," The Christchurch Call, July 3, 2024, https://www.christchurchcall.org/.
[44] "Forum on Information & Democracy," Forum Information & Democracy, accessed July 18, 2024, https://informationdemocracy.org/principles/.

International Declaration on Information and Democracy, a framework developed by notable figures from media, academia, and civil society.

Protecting Children Online: The WePROTECT Global Alliance[45] brings together 150+ governments, tech companies, and civil society organisations to combat online child exploitation and abuse through coordinated actions and shared resources. The Children Online Protection Lab[46] is dedicated to ensuring the safety of children in digital spaces by facilitating the development and implementation of essential tools and guidelines among stakeholders for online protection.

Detecting Inauthentic Content: In anticipation of potential threats to the integrity of the 2024 elections, various tech companies have formed the Tech Accord to Combat Deceptive Use of AI.[47] This initiative aims to detect and prevent the spread of inauthentic content generated by AI, ensuring that election information remains accurate and trustworthy.

6. **There are still critical matters pending resolution**, awaiting the response of an international framework:

- **Data transfer** is vital for AI deployment as it provides access to large and diverse datasets, but restrictions on cross-border data transfers (CBDT) are on the rise, as illustrated by a recent set of CERRE papers[48]. Various regions are attempting to reconcile three competing interests: privacy and data protection, digital trade, and data sovereignty, with the latter emerging as a multifaceted concept used to achieve multiple regulatory objectives, including strategic economic autonomy, cyber resilience, and national security.

  The landscape is slowly consolidating without converging. The European Commission has recently adopted the EU-US Data Privacy Framework and validated 11 pre-GDPR adequacy decisions. China has released new provisions on CBDT, indicating a willingness to relax its CBDT regime. India, with its new data protection law, has opted for a blacklist approach to CBDT, although sectoral constraints remain in place. Brazil, now appearing as a serious candidate for EU adequacy, has recently announced a draft regulation related to international data transfers. Meanwhile, the WTO's new digital trade agreement, despite its limited focus on data flows, sets foundational rules for digital trade facilitation among 80+ member countries. The current lack of binding provisions on cross-border data movement reflects the ongoing global challenges in harmonising data governance.

  G20 leaders' statements have repeatedly emphasised the importance of cross-border data flows, particularly since the Osaka Declaration of 2019, which introduced the concept of Data Free Flow with Trust (DFFT). But addressing regulatory challenges related to data protection would necessarily involve building mechanisms that can bridge the jurisdiction of the data

[45] "WeProtect Global Alliance," accessed July 18, 2024, https://www.weprotect.org/alliance/.
[46] "Children Online Protection Lab Charter," elysee.fr, November 8, 2023, https://www.elysee.fr/en/emmanuel-macron/2022/11/10/laboratory-for-childhood-protection-online-charter.
[47] "A Tech Accord to Combat Deceptive Use of AI in 2024 Elections," AI Elections accord, accessed July 18, 2024, https://www.aielectionsaccord.com/.
[48] Sophie, Stalla-Bourdillon, Global Governance of Cross-Border Data Flows, September 2024, https://cerre.eu/publications/global-governance-of-cross-border-data-flows/

exporter and importer, respectively, without compromising national approaches. They would likely require the development of a new, multi-layered approach to CBDT.

- A growing debate touches upon **intellectual property** and is centred around the increasing prevalence of generative AI in creative industries, which raises complex legal questions.

  The core issue is the lack of clarity in copyright laws when applied to AI-generated works, particularly concerning ownership and copyright infringement.[49] As AI systems create content that rivals human creativity, the legal system struggles to define the boundaries of IP rights.[50] Notably, the use of unlicensed content in AI training data and the provenance of AI-generated content are contentious points. Courts are actively trying to navigate these uncertainties, with several cases already filed, aiming to establish precedents for future applications of IP laws to generative AI.

- Another key dimension is the **tension between open-source and proprietary models of AI development.**

  Open-source AI has historically fostered innovation by allowing researchers and developers to build on each other's work, share best practices, and accelerate scientific discovery.[51] It has also enabled greater scrutiny and verification of the reliability of AI systems, providing opportunities for feedback and improvement.[52] Open sourcing AI would accelerate the democratised access to AI technology, empowering small businesses, civil society, and under-resourced regions to benefit from and contribute to AI development.[53] Moreover, if future AI systems are envisioned as a common utility or a personal interface to the world, this would support a rationale for decentralised, diverse, and broadly accessible AI models, rather than proprietary systems in which AI controlled by a few.

  Open-source also faces critics. Safety concerns cannot be dismissed, as open sourcing could facilitate the misuse or abuse of AI by malicious actors, such as hackers, criminals, or rogue states.[54] The availability of powerful open-source AI tools such as deepfake generators, facial recognition software, or cyberattack methods, could pose serious threats. Geopolitical pressures are additionally exerted, as some major players are envisioning broadly restrictive limits on open innovation on national security grounds. Export controls on AI technology to prevent strategic advancements from falling into the hands of rival states would also inevitably hamper global collaboration and knowledge exchange.

  The safety concerns raised regarding open-source AI are legitimate, but they do not justify a blanket restriction. Responsible open sourcing guidelines and principles could be developed and followed – including measures such as risk assessments, documentation, licensing,

[49] Vincenzo Iaia, "To Be, or Not to Be … Original Under Copyright Law, That Is (One of) the Main Questions Concerning AI-Produced Works," GRUR International 71, no. 9 (September 26, 2022): 793–812, https://doi.org/10.1093/grurint/ikac087.
[50] Alesia Zhuk, "Navigating the Legal Landscape of AI Copyright: A Comparative Analysis of EU, US, and Chinese Approaches," AI and Ethics, May 30, 2023, https://doi.org/10.1007/s43681-023-00299-0.
[51] Arthur Spirling, "Why Open-Source Generative AI Models Are an Ethical Way Forward for Science," Nature 616, no. 7957 (April 18, 2023): 413–413, https://doi.org/10.1038/d41586-023-01295-4.
[52] Manuel Hoffmann, Frank Nagle, and Yanuo Zhou, "The Value of Open Source Software," SSRN Scholarly Paper (Rochester, NY, January 1, 2024), https://doi.org/10.2139/ssrn.4693148.
[53] Nathan Benaich, and Alex Chalmers, "The Case for Open Source AI," April 18, 2024, https://press.airstreet.com/p/the-case-for-open-source-ai.
[54] Francisco Eiras et al., "Near to Mid-Term Risks and Opportunities of Open-Source Generative AI" (arXiv, May 24, 2024), http://arxiv.org/abs/2404.17047.

accountability, and oversight – balancing innovation with safety. Export controls on widely used software technologies, for their part, have often proved to be counterproductive; the US restrictions on encryption in the 1990s were ultimately rolled back, as they hindered the development of secure online communications and commerce, while doing little to prevent adversaries from accessing encryption tools.

A preferable and plausible scenario would be to prevent a sharp division between open-source and proprietary AI, and instead have degrees of openness. Different levels of openness may be appropriate for different types of AI applications, depending on their potential benefits and risks, as well as the intended users and audiences. Both models can operate side by side, and developers should follow and benefit from standards and governance initiatives that are customised to how they publish their models, depending on various factors. A fine-grained comprehension of the different degrees of openness and how they affect safety and innovation is required, as well as policies that would hinder global cooperation and innovation.

7. **Three functions are required at a global level**, as they are essential for ensuring that AI serves the common good and does not exacerbate existing tensions.

- **Shared Assessment of AI Advancements and Risks**

  There is an acute need for a global framework to synthesise and disseminate the latest scientific understanding of AI's capabilities and risks, and provide a central, authoritative source of scientific knowledge to inform global decision-making.

  This framework would regularly produce expert-led assessments, which are updated frequently to keep pace with AI's rapid innovation. Maintaining policy neutrality is crucial to foster consensus and trust among a diverse set of stakeholders, ensuring that the assessments are seen as unbiased and authoritative. The participation of experts from a wide array of disciplines and countries, including low- and middle-income nations, would ensure a comprehensive and inclusive global perspective on AI. The initiative to commission an international report on AI by the UK AI Safety Summit, led by Yoshua Bengio, is a preliminary step in this direction.[55]

  Several models inspire this approach. The Intergovernmental Panel on Climate Change's (IPCC) structure is the most cited, which involves thousands of scientists producing comprehensive reports, provides a blueprint, though AI's swift evolution necessitates more frequent updates.[56] A streamlined model, potentially with separate panels for different AI aspects and a real-time horizon-scanning function, could address these needs, ensuring the timely synthesis and dissemination of AI advancements and risks.

  Other models exist though. The Montreal Protocol which provides for separate panels for each of the various issues at stake enhances efficiency and specialisation. A central registry for up-to-date AI developments and a horizon-scanning function to alert the global community to emerging risks could constitute essential components of the global framework. Or an agency

[55] Bengio Yohsua et al., "International Scientific Report on the Safety of Advanced AI" (Department for Science, Innovation and Technology, May 2024), https://hal.science/hal-04612963.
[56] Joseph Bak-Coleman et al., "Create an IPCC-like Body to Harness Benefits and Combat Harms of Digital Tech," Nature 617, no. 7961 (May 2023): 462–64, https://doi.org/10.1038/d41586-023-01606-9.

model, mirroring the role of the World Health Organization (WHO) in global health (often referred to as "the health agency of countries that don't have health agency") could provide the world with a common baseline of necessary understanding and scientific assessment for informed decision-making.

- **Equitable Access and Benefits**

  AI's transformative potential poses a risk of deepening the digital divide, especially for low- and middle-income countries (LMICs). To prevent this, specific development instruments are necessary to address market failure and promote equitable access and benefits. These instruments include capacity-building efforts, technology transfer, and financial support.

  LMICs need assistance in building their AI capabilities. This includes providing access to computational resources, data, and existing AI models. Public-private partnerships in global health offer valuable lessons. The Global Alliance for Vaccines and Immunizations (Gavi) and the Global Fund to Fight AIDS, Tuberculosis, and Malaria demonstrate how international collaboration can unlock financing and make innovative technologies accessible to LMICs. A similar approach for AI could play a pivotal role in developing AI models, applications, and human capital, tailored to the needs of developing countries. This model would involve a multi-stakeholder framework, incorporating public-private partnerships, international organisations, and philanthropic foundations.

- **Safety and Risks Mitigation**

  Ensuring the safe development and use of AI technologies involves creating taxonomies of risks, establishing benchmarks and evaluations, and building processes for external validation and standardisation. These may differ across jurisdictions, barring a specific effort to the contrary. For example, the Bletchley Park Summit initiated discussions on this front.

  Establishing international standards for AI development and deployment, akin to those of the International Organization for Standardization (ISO) and forming a global network of AI safety institutes to monitor compliance and validate AI systems, would enhance safety. This network would develop and standardise risk assessments, benchmarks, and evaluation protocols. A comprehensive classification system for AI risks, from algorithmic bias to existential threats, alongside global benchmarks for AI safety, transparency, and accountability, forming the backbone of the safety and risk mitigation effort, is ambitious but achievable.

  Long-term control and governance of AI technologies requires robust oversight mechanisms to ensure the technologies remain under human control. The process should be rigorous yet flexible, allowing for innovation.

  Other models have been proposed, but these seem far-fetched and some are likely to have unintended side-effects. An organisation similar to the International Atomic Energy Agency (IAEA), which would oversee the development of advanced AI – enforcing safety standards and conducting audits – may be premature given the currently evolving state of the technology.[57] A multilateral export control regime could oversee the export of key AI technologies to prevent misuse. However, a regime modelled on the Nuclear Suppliers Group

[57] Seokki Cha, "Towards an International Regulatory Framework for AI Safety: Lessons from the IAEA's Nuclear Safety Regulations," Humanities and Social Sciences Communications 11, no. 1 (April 12, 2024): 1–13, https://doi.org/10.1057/s41599-024-03017-1.

could risk stifling innovation or hindering access to essential uses. A Conditional Access Framework, inspired by the Nuclear Non-Proliferation Treaty (NPT)'s provisions on peaceful uses of nuclear energy, which would ensure that access to AI technology is contingent on adherence to safety and ethical standards, could be a potential option. However, such conditionality for an essential technology would likely face stark resistance from recipient countries.

8. **Control of AI use cases for offensive purposes is the last crucial aspect in ensuring the safety of AI.** Discussions on this would occur on discrete, confidential tracks. These discussions would aim to limit the development, possession, and use of AI weapons, where the threat to human security is deemed to be unacceptable.[58] Depending on the level of risk and the degree of consensus among participating states, different models of arms control regimes could be applied.

   One potential model is a limitation regime, which sets quantitative or qualitative restrictions on certain types of AI weapons, e.g., autonomous lethal weapons or cyberattack tools. This would aim to reduce the risk of arms races, escalation, or destabilisation, while allowing for legitimate objectives in terms of defence and deterrence. A limitation regime could be modelled after treaties like the Strategic Arms Limitation Talks (SALT) or the Strategic Arms Reduction Treaty (START), which imposed ceilings and verification mechanisms on nuclear weapons.

   This would likely be associated by non-proliferation, i.e., preventing the spread of AI weapons to additional actors, especially those who might use them irresponsibly or maliciously. Such regimes aim at preserving the existing balance of power and prevent rogue states, terrorists, or criminals from acquiring dangerous AI capabilities. A non-proliferation mechanism may eventually be modelled after treaties like the NPT, establishing a framework of obligations and inspections for nuclear weapons states and non-nuclear weapons states: it would focus on AI-powered weapons specifically.

   A third possible model could aim at specific renunciation or interdiction, prohibiting the development, possession, and use of certain AI weapons altogether, and require their elimination or confiscation. Though rare in the international system, complete eradication of specific weapons based on their classification as inherently immoral, illegal, or unacceptable, exists: the Chemical Weapons Convention (CWC) bans the production and stockpiling of chemical weapons and mandates their destruction. A renunciation or interdiction regime on very specific AI weapons could be modelled on such mechanisms.

   Choosing the most appropriate model for controlling offensive AI capabilities depends on political viability and technical feasibility – alas, surpassing moral desirability. The effectiveness of any arms control regime depends on the cooperation and compliance of all relevant actors. The development of such regimes would require extensive dialogue, negotiation, and coordination among states.

[58] Henry A. Kissinger and Graham Allison, "The Path to AI Arms Control," Foreign Affairs, October 13, 2023, https://www.foreignaffairs.com/united-states/henry-kissinger-path-artificial-intelligence-arms-control.

For now, the most likely steps would be processes for supercomputer transparency and rules of engagement in AI development. Such a framework would require participating entities to disclose their AI capabilities and development practices, fostering an environment of mutual understanding that could eventually lead to more comprehensive approaches.

4. Conclusion

The convergence of AI with global governance is an extraordinary opportunity and a profound challenge. AI has the capacity to revolutionise numerous sectors including healthcare, education, and industry, heralding a new era. But to fully realise these benefits, deliberate and strategic policies must be implemented to foster innovation, ensure safety, and promote fairness.

Rapid advancements in AI necessitate a global governance framework that keeps pace with technological developments. Such a framework should foster innovation by providing incentives for research and development, while also addressing the potential risks associated with AI. At the same time, the potential for AI to exacerbate inequalities both within and between countries must be addressed through targeted policies. This includes measures to support the reskilling and upskilling of workers displaced by automation, as well as the provision of social safety nets.

Safe deployment of AI systems is paramount, as AI systems become more integrated in daily life. AI systems must be trustworthy, transparent, and accountable, and aligned with human values and interests. Establishing international standards for AI safety, transparency, and accountability, akin to the ISO, can enhance global trust. This includes rigorous testing, transparent methodologies, and the implementation of external validation processes. Effective governance of AI also involves addressing ethical implications and societal impacts. This includes protecting human rights, ensuring data privacy, and preventing the misuse of AI technologies. Frameworks established by the G7, OECD, and UNESCO offer varying degrees of universal applicability.

Fair access should ensure that the benefits of AI should be distributed equitably, preventing the technology from widening the existing global digital divide. Low- and middle-income countries must be supported through capacity-building initiatives, technology transfers, and financial aid. Public-private partnerships, like those in global health, could play a pivotal role in making AI accessible to all. Ensuring that the benefits of AI are widely shared, requires international collaboration and the adoption of policies that promote inclusivity.

The geopolitical dynamics surrounding AI present a significant challenge to global governance. A single global AI governance regime is unrealistic due to varying national interests and cultural differences. Instead, a regime complex — a network of overlapping institutions and agreements — can accommodate diverse preferences and promote coexistence. Akin to global trade systems, a network of bilateral and regional agreements can facilitate technology exchange, data sharing, and ethical standards, serving as intermediaries between local and global frameworks and creating a resilient web of governance that adapts to AI's rapid evolution.

AI governance should prevent sharp divisions between open-source and proprietary models. Instead, different levels of openness should be tailored to AI applications' benefits and risks. Degrees of openness, rather than strict dichotomies, will foster innovation while ensuring safety. This would also help bridge the technological gap between the Global North and the Global South and ensure that all countries can benefit from AI advancements.

Three critical functions are essential for effective AI governance: a shared assessment of AI advancements and risks, equitable access and benefits, and safety and risk mitigation. Establishing a

global framework for assessing AI's capabilities and risks can provide authoritative and unbiased scientific knowledge to inform policy decisions. Ensuring equitable access to AI technology requires specific development instruments to support LMICs. Mitigating the risks associated with AI involves the creation of international standards and the establishment of a global network of AI safety institutes.

Finally, controlling offensive AI capabilities requires international cooperation. Limitation, non-proliferation, and renunciation models could be applied to specific AI weapons, ensuring they do not pose excessive threats to human security.

Our collective future hinges on balancing rapid innovation with ethical integrity and equitable access. AI's promise is vast. It comes with responsibilities. Embracing a coexistence of diverse approaches will be our best defence against fragmentation in the uncharted waters ahead. Ultimately, our goal is not to manage AI but to harness its potential for the collective good of humanity.

cerre Centre on Regulation in Europe

Avenue Louise 475 (box 10)
1050 Brussels, Belgium
+32 2 230 83 60
info@cerre.eu
www.cerre.eu

Centre on Regulation in Europe (CERRE)
CERRE Think Tank