Stockholm University

SCGG Policy Brief
2026:1 (September)

Making Global AI Governance Work: Why We Need Consolidated, Interoperable, and Binding Rules on AI Now

Johannes Geith

Abstract

Artificial Intelligence (AI) is widely considered the most significant technological development of the 21st century. Widespread adoption of increasingly powerful AI promises everything from gains in productivity that will propel economic prosperity, to advances in medical research and disease detection. At the same time, AI carries real risks. Governments, international organizations, and firms are racing to capture AI's benefits while mitigating its dangers. Given that many AI systems are developed in cross-national processes and produce transboundary effects, most agree that effective AI governance will have a global dimension. This policy brief has three main aims. First, it discusses the governance challenges posed by AI and identifies why some of these require international coordination. Second, it surveys the landscape of global AI governance efforts and identifies its key shortcomings. Third, it concludes by making recommendations to address these weaknesses in existing AI governance.

AI: A (Global) Governance Challenge?

Generative AI—a type of AI that, for example, creates original text, images, audio, and code—is now being adopted at a faster rate than when the Internet first emerged. The Internet once raised hopes that it would offer a space where knowledge, ideas, and democracy would thrive. These hopes only partly materialized, and its spread also brought entirely new dangers and policy challenges, from disinformation and hate speech, to the rise of mass collection, tracking, and commercial use of personal data. AI similarly gives rise to a wide array of hopes, including, but not limited to, hope that it might lead to productivity gains, economic prosperity, and medical breakthroughs.

At the same time, and similar to the Internet, AI poses significant risks. Both unintentional technical failures and the intentional misuse of AI can have highly problematic consequences. AI can, for example produce deepfakes, which are AI-generated but extremely realistic pictures, videos, or audio clips, which have been linked to undermining trust in elections and pose serious dangers for liberal democracies. Widespread concerns also exist about fairness—as AI-induced automation may lead to mass-layoffs—and accountability and human rights violations—as autonomous weapons systems may alter the ways that armed conflicts are fought today. Citizens are increasingly opposing data center construction, which are important to train AI, citing concerns over massive water consumption and environmental impacts. Finally, at the far end of the risk spectrum lies Artificial General Intelligence (AGI), AI that matches or even surpasses human capabilities. Experts are increasingly worried that AGI could pose a threat to humanity's survival.

To reap the benefits of AI, while mitigating its risks, there have been increasing efforts to establish appropriate guardrails and put in place regulations on AI. Yet fundamental conflicts persist over the *who*, *what*, and *how* of AI regulation. The question of who should regulate AI is divided across multiple fault lines: national versus international levels of governance, and public versus private regulation. There are also disputes over what should be regulated. Conflicts persist on whether AI is best governed through broad general rules that apply to all AI, or specific rules tailored to particular AI applications and uses (e.g., healthcare or finance). The question of how to regulate AI involves fundamental choices about which policy instruments are most appropriate, for instance questions about whether these should be binding or non-binding—that is, legally enforceable or merely voluntary.

Some AI governance challenges are adequately addressed nationally. However, others—in light of AI's transnational development and cross-border effects—clearly require international coordination and collaboration. These will be the focus of this policy brief. Consider the case of AI weaponry. Autonomous weapons systems are weapons systems with certain AI capabilities that can identify and eliminate targets without any human operator. These therefore remove human judgment from decisions that are by their nature life and death. The implications of the use of these AI-controlled weapons systems are likely to extend beyond national borders, carry profound consequences for military strategy between countries, and most importantly, raise fundamental ethical questions about accountability. Arms control regimes have often been the answer to dangerous military applications that raise serious concerns over ethics, legality, stability, and safety. AI weaponry, by most accounts, checks all these boxes. The Convention on Certain Conventional Weapons (CCW) provides a framework within which new regulations on the use of AI weaponry could be established. However, negotiations over the regulation of AI weaponry have largely stalled and results are limited to non-binding guiding principles. The lack of progress establishing regulations is not unique to military AI. This raises two key questions: What is the state of global AI governance today? What accounts for this current state?

The State of Global AI Governance: Heterogeneous Actors, Institutional Fragmentation, and Non-Binding Rules

Global AI governance today is characterized by three defining trends. First, the actors involved in AI governance are heterogeneous. Core actors are states, international organizations, and civil society organizations, but powerful AI companies, by virtue of their technical expertise, have become the ones effectively setting the rules. The result is a governance landscape where institutions like the UN and the Council of Europe share the field with the frameworks which powerful AI companies, such as Google, OpenAI, and Anthropic, have established to regulate their own products.

Second, governance efforts are fragmented across many separate institutions. They involve a mix of old and new institutions whose responsibilities often overlap or clash. For example, the OECD, the Global Partnership on AI, and the EU all operate within the same policy space, with partly overlapping membership and functions. These institutions often do not clearly divide responsibilities. This leads to duplicated efforts, where several institutions perform the same task, while other governance gaps persist and remain unaddressed.

Third, most global rules on AI to date are non-binding. They typically involve voluntary guidelines rather than enforceable laws. Even landmark AI policy efforts, such as the 2023 Bletchley Declaration on AI safety, carry no legal force. Meanwhile, the EU AI Act—despite its ambitious risk-classification framework—remains limited in its geographic scope and has faced considerable headwinds since its adoption in 2024.

These trends can be attributed to many factors, two of which stand out. First, the involved actors hold sharply divergent preferences. This includes differences in the views of the same kind of actor. For instance, some states and corporations prioritize enabling innovation and capturing AI's economic potential, while others emphasize the safety risks that AI may produce. These differences also extend across actor groups, with states, international organizations, corporations, and civil society actors each bringing fundamentally different interests, priorities, and visions of what AI governance should achieve.

Second, there remains broader uncertainty about AI itself. AI is still evolving rapidly. Its capabilities and risks remain difficult to predict, and actors have to navigate difficult trade-offs. They need to balance between reaping the gains from cooperation and securing competitive advantages over a strategically significant technology.

Fragmentation in global governance is hardly new or unique to AI. Across transnational policy domains like trade or climate, the global governance landscape has become progressively more complex. What sets AI apart, however, is the exceptional prominence of private actors. Another distinguishing characteristic of AI governance is the speed at which both existing institutions and novel arrangements have emerged (see Figure 1).

Implications and Recommendations: Consolidation, Interoperability, and Binding Rules

The diversity of governance challenges posed by AI is unlikely to be met by a one-size-fits-all approach. Instead, effective responses will require varied forms of cooperation. Yet current efforts are widely regarded as inadequate. This policy brief has highlighted three trends that help explain why. Concrete steps can nonetheless be taken to address these shortcomings. Here, three such steps are highlighted.

First, global AI governance needs a clearer division of labor between institutions. Policymakers should identify the most consequential governance challenges requiring global action and concentrate the relevant governance functions accordingly. They should make use of existing institutions where possible and devel-

op new ones where necessary. Critically, such consolidation should be guided by what actually needs to be done—not institutional rivalry or political self-interest.

Second, regulators should prioritize interoperability when designing AI rules, so that rules in different countries and institutions do not conflict or duplicate efforts. Rather than pursuing a single unified regime, efforts should focus on making existing regulatory frameworks more compatible with one another. The recently launched UN working group as part of the Global Dialogue on AI Governance is a step in the right direction.

Third, state parties and signatories of the Bletchley Declaration should commit to negotiating binding obligations that champion AI safety for the most severe risks immediately. The current reliance on soft law and voluntary instruments reflects the uncertainties surrounding AI. However, there is now broad agreement that AI technology is mature enough to warrant firmer rules. It is imperative that policymakers agree on what are seen to be the core risks of AI, something which at the moment remains deeply contested. Once they have done this, they also need to commit to establishing binding obligations to address these risks.

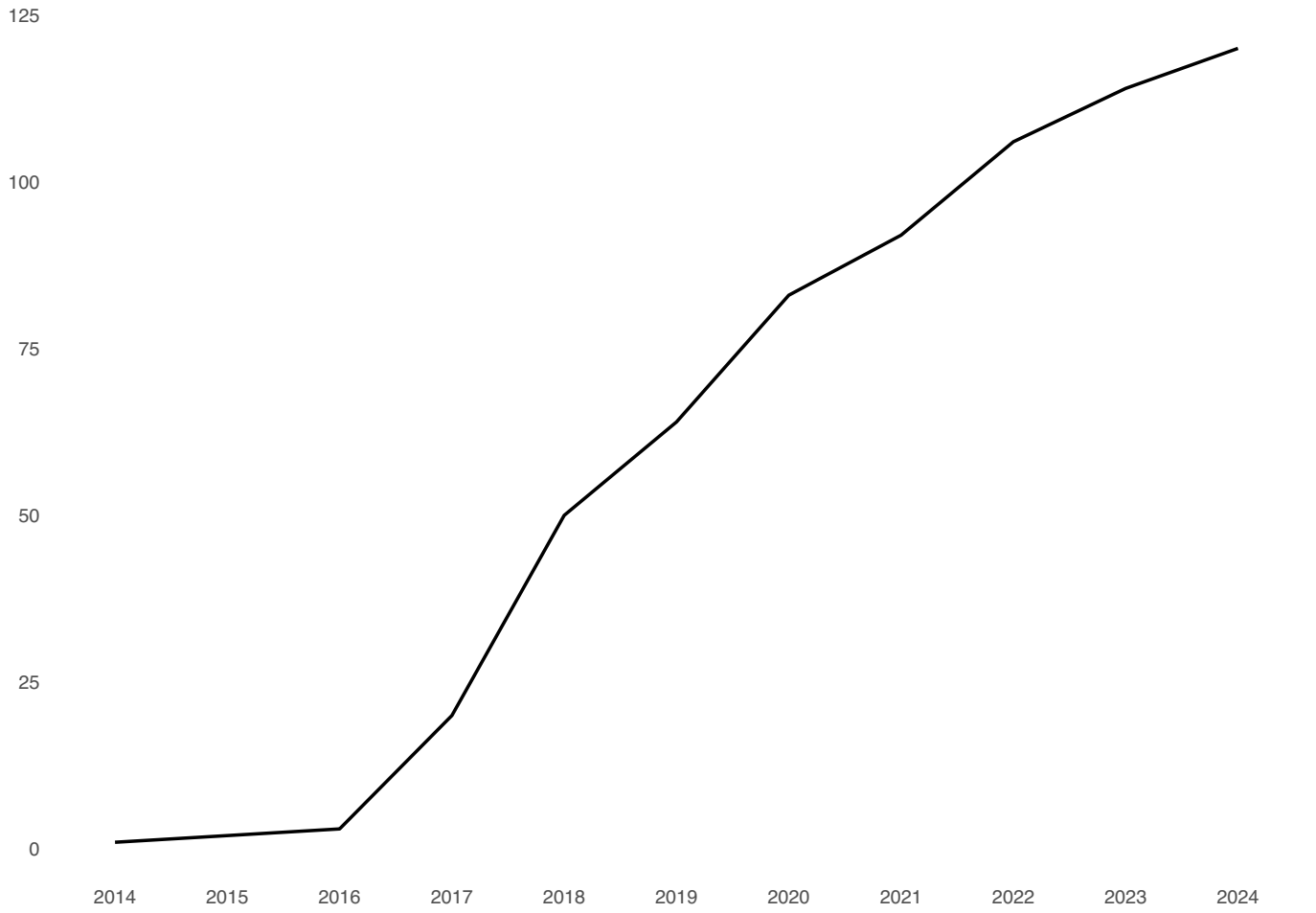


Figure 1: Cumulative count of international AI institutions, 2014-2024. Source: Geith, Johannes, Magnus Lundgren, and Jonas Tallberg (2026). The Emerging Regime Complex for Artificial Intelligence. Global Studies Quarterly: Forthcoming.

Takeaways

- AI is being adopted faster than the Internet was, bringing parallel promise and peril. AI offers hopes of productivity gains and economic prosperity but also poses serious risks—including existential threats through Artificial General Intelligence (AGI).
- Global AI governance is defined by three trends: heterogeneity of actors, institutional fragmentation, and non-binding rules.
- Two main factors are especially prominent in explaining the current state of affairs: divergent preferences and uncertainty about AI itself.
- Three concrete steps could be taken to improve AI governance: establishing a clearer division of labor and consolidate efforts, greater interoperability, and binding commitments on the most serious risks.

Further Reading

Furendal, Markus and Magnus Lundgren (2026). *Handbook on the Global Governance of Artificial Intelligence*. Edward Elgar Publishing.

Geith, Johannes (2026). *The Global Governance of Artificial Intelligence: Actor Preferences, Bargaining Dynamics, and Institutional Design*. Doctoral Dissertation. Stockholm Studies in Politics 207.

Yudkowsky, Eliezer and Nate Soares (2025). *If Anyone Builds It, Everyone Dies: Why Superhuman AI Would Kill Us All*. Little, Brown and Company.

About the SCGG Policy Brief Series

The Stockholm Center on Global Governance (SCGG) is an international, interdisciplinary research center at Stockholm University. The SCGG Policy Brief Series serves to bridge academic research and policy practice.

Chief Editor
Mark Klamberg

Editorial Committee
Paul Levin
Sally Longworth
Nicholas Olczak
Mark Rhinard

About the Author

Johannes Geith is a Postdoctoral Fellow at the Department of Political Science at the Massachusetts Institute of Technology (MIT). He obtained his PhD in June 2026 at the Department of Political Science at Stockholm University and is affiliated with the Stockholm Center on Global Governance. His research focuses on the politics of artificial intelligence; his wider research interests include global governance and international political economy. His work has been published or is forthcoming with *International Studies Review, Business & Politics, Global Policy*, and *Global Studies Quarterly*.

scgg@su.se
www.su.se/scgg

Please register for future SCGG policy briefs here.

The views and opinions expressed in this policy brief are those of the author and do not necessarily reflect the positions of the SCGG.