Complexity and the Global Governance of AI

Gordon LaForge[1], Anne-Marie Slaughter[1], Simon Levin[2], Adam Day[3], Allison Stanger[4], Ann Kinzig[5], Stephanie Forrest[5], Bruce Schneier[6], Cristopher Moore[7], Kevin O'Neil[8], Moshe Vardi[9], Nazli Choucri[10], Robert Axelrod[11], Sihao Huang[12], Steve Crocker[13], Tina Eliassi-Rad[14], Nick Silitch[15], Merle Weidt[16]

[1]New America, [2]Princeton University, [3]UN University Centre for Policy Research, [4]Middlebury College, [5]Arizona State University, [6]Harvard Kennedy School, [7]Santa Fe Institute, [8]Rockefeller Foundation, [9]Rice University, [10]Massachusetts Institute of Technology, [11]University of Michigan, [12]University of Oxford, [13]Edgemoor Research Institute, [14]Northeastern University, [15]Prudential (*retired*), [16]University of Denver

Abstract

In the coming years, advanced artificial intelligence (AI) systems are expected to bring significant benefits and risks for humanity. Many governments, companies, researchers, and civil society organizations are proposing, and in some cases, building global governance frameworks and institutions to promote AI safety and beneficial development. Complexity thinking, a way of viewing the world not just as discrete parts at the macro level but also in terms of bottom-up and interactive complex adaptive systems, can be a useful intellectual and scientific lens for shaping these endeavors. This paper details how insights from the science and theory of complexity can aid understanding of the challenges posed by AI and its potential impacts on society. Given the characteristics of complex adaptive systems, the paper recommends that global AI governance be based on providing a fit, adaptive response system that mitigates harmful outcomes of AI and enables positive aspects to flourish. The paper proposes components of such a system in three areas: access and power, international relations and global stability; and accountability and liability.

I. Background

For all its potential benefits, advanced artificial intelligence (AI) presents serious and uncertain threats to global stability. Disinformation, cyberattacks, labor displacement, inequality, autonomous weapons systems – managing these risks and others is an urgent priority for global governance.[1] Several bilateral, regional, and international processes are underway that aim to facilitate international cooperation and develop shared governance frameworks for AI, though it

[1] Defined by Thomas G. Weiss as "formal and informal processes to manage issues of global concern."

is early days still with policy makers working to understand AI and its likely effects on society, geopolitics, and humanity.[2]

Complexity thinking, a way of viewing the world not just as discrete parts at the macro level but also in terms of bottom-up and interactive complex adaptive systems, can help us understand generative AI and how to govern it. In a complex adaptive system:

- Heterogeneous collections of individual agents interact locally and evolve in surprising, unpredictable ways due to those interactions.[3]
- Fueled by multiple feedback loops, systems self-organize and develop emergent patterns, coalescing around steady states, limit cycles, or more complicated asymptotic (limit-approaching) behaviors called "attractors."
- Change is non-linear, and a system can suddenly shift from one attractor basin to another.
- Conflict inevitably emerges between scales, as the interests of individuals, collectives, and society come into tension.
- The behaviors and outcomes of a system are emergent, more and different than the sum of its constituent parts.[4]

Most biological, technological, and human systems are complex adaptive systems. The global economy is one example, as is the interconnected bio-eco-socio-technosphere.[5]

When it comes to global AI governance, complexity is a useful heuristic. The background condition of continual change makes complexity thinking feel familiar and even welcoming for decision-makers and practitioners struggling to shift from prediction and planning to uncertainty and strategic adaptation. Moreover, the assumption of complexity recognizes that a fundamental challenge for governance is in dealing with the relationships among multiple scales of space, time, and organization. Micro processes and interactions at finer scales can maintain or undermine the resilience and robustness of macro structures, especially the public goods and common-pool resources that are key to sustainability.

The question remains, however, whether complexity thinking can also point us toward specific and applicable governance strategies, tools, and mechanisms. If systems shift among basins of

---

[2] Several thoughtful, robust proposals already exist, notably: the UN High Level AI Advisory Body's Interim Report: Governing AI for Humanity (December 2023); the DeepMind research paper by Lewis Ho et al. International Institutions for Advanced AI (July 2023); and the Artificial Intelligence & Equality Initiative's Framework for the International Governance of AI (July 2023).

[3] For more on complex adaptive systems and their properties see P.W. Anderson, "More Is Different," Science, New Series, Vol. 177, No. 4047. (Aug. 4, 1972), pp. 393-396; Simon Levin, "The architecture of robustness," in *Global Challenges, Governance, and Complexity: Applications and Frontiers* (2019), pp. 16-23; Adam Day, *The Forever Crisis*: *Adaptive Global Governance for an Era of Accelerating Complexity*, Routledge (2024).

[4] P.W. Anderson, "More Is Different," *Science*, New Series, Vol. 177, No. 4047. (August 1972), pp. 393-396

[5] Kenneth J. Arrow, Paul R. Ehrlich, and Simon A. Levin, "Some Perspectives on Linked Ecosystems and Socioeconomic Systems," in *Environment and Development Economics: Essays in Honour of Sir Partha Dasgupta*, Oxford University Press (2014).

attraction, can we create new attractors and influence dynamics to cause a system to flip from a harmful attractor to a more desirable one? If so, how? And if different scales of the system tend to be in conflict, how can we develop rules that enable us to manage a global commons across conflicts of interest? Scholars and practitioners are already grappling with these questions.[6] This paper seeks to build on the work already done and from it derive proposals that may encourage the safe and equitable development of AI globally.

II. Key Issues and Challenges for Global AI Governance

Several features of AI and the socioeconomic systems surrounding it make the technology distinctly difficult to govern. These include:

The Timescale Mismatch. Changes in law, regulation, and policy are incremental, whereas AI systems are developing at an exponential rate.[7] To give one indicator, training computation, which is one of three factors that determine the capability of an AI system, has roughly doubled every six months since 2010.[8] Governance institutions and processes struggle to keep up. The result is a "Collingridge Dilemma," whereby addressing the full risks of AI may only become a policy priority after society has become too reliant on the technology to change course.[9] This mismatch of scales is a prototypical feature of complex adaptive systems, and methods to deal with that conflict are part and parcel of the theoretical foundations of the subject.

Unknown Unknowns.[10] The advance of AI entails irreducible uncertainty about the path of the technology itself and its potential socioeconomic and political impacts. Researchers doubt whether generative AI systems have actually displayed emergent abilities, as some have claimed[11]; still, to an even greater degree than with other complex risks such as climate change, the world lacks a reliable forecast of even many of the near-term possibilities and risks of AI. Part of the uncertainty is that advanced AI systems are opaque (possessing too many parameters for a single human to easily understand their inner workings); surprising (tending to increase

---

[6] See Adam Day, *The Forever Crisis*: *Adaptive Global Governance for an Era of Accelerating Complexity*, Routledge (2024), in which he proposes five essential components of adaptive governance informed by complexity thinking: (1) map the system; (2) design fit, flexible institutions; (3) include many stakeholders; (4) iterate continuously based on feedback; (5) focus on attractors and underlying rules and structures. See also the work of the Cascade Institute, https://cascadeinstitute.org; the Santa Fe Institute, https://www.santafe.edu; and the Complexity Science Hub, https://csh.ac.at;

[7] See Thomas Hale, *Long Problems: Climate Change and the Challenge of Governing Across Time*, Princeton University Press (2024).

[8] https://ourworldindata.org/brief-history-of-ai.

[9] https://demoshelsinki.fi/2022/02/15/what-is-the-collingridge-dilemma-tech-policy/

[10] Ralph Gomory, "The Known, the Unknown, and the Unknowable," *Scientific American* (June 1995), https://www.scientificamerican.com/article/the-known-the-unknown-and-the-unkno/.

[11] For two opposing perspectives on whether large language models display emergent abilities see Jason Wei, et al., "Emergent Abilities of Large Language Models," *Transactions on Machine Learning Research* (August 2022), https://arxiv.org/pdf/2206.07682.pdf, and Ryan Schaeffer, Brando Miranda, and Sanmi Koyejo, "Are Emergent Abilities of Large Language Models a Mirage?," 37th Conference on Neural Information Processing Systems (NeurIPS 2023), https://proceedings.neurips.cc/paper_files/paper/2023/file/adc98a266f45005c403b8311ca7e8bd7-Paper-Conference.pdf.

capacities in unanticipated ways); and easy to misuse (especially when it comes to general-purpose AI systems). Uncertainty is not a reason to dismiss traditional regulatory approaches, but it does increase the challenge, in part because humans tend to overestimate the short-term impacts of technologies and underestimate their longer-term effects.[12] Policymakers may panic about a new AI function and overreact, regulating it in a way that locks in some of what turn out to be its more dangerous aspects in the longer term.

Concentrations of Corporate Power. AI infrastructure and development are highly concentrated in private companies, and a handful of American and Chinese companies at that. The most powerful models are produced by the private sector, and industry R&D is far better financed and advanced than that in universities, governments, and other public institutions. Not only does concentration of corporate power over AI risk limiting the applications of systems to maximizing shareholder returns and other capitalist imperatives, but it also means that as a practical matter companies will play a major role in the global governance of AI, either intentionally or by default. The conflicts of interests among agents at different scales – corporations at one level, governments at another – create challenges that are characteristic of the governance of all complex adaptive systems. State-centric approaches to global governance continue to struggle with this reality – especially when the corporations involved are more powerful than the majority of the world's governments, as is the case with AI.

Inequality. Already, the digital divide is a major source of global inequality. AI stands to turn it into a digital chasm. In the near-term, the concentration of AI system development in the rich world will result in most of the economic gains from the technology accruing to those countries. In the long-term, AI is likely to become individualized, and those who have access to the technology will lead different, more prosperous lives than those who lack access to it. Feedback loops and preferential attachment, a feature of networks that explains why the rich get richer, mean the poor will fall ever further behind.[13] Thus, at the heart of AI development is an emergent tension between equity — between North and South, and between governments and their citizens — and the tendency toward concentration referred to above.

Widely Distributed Harms. AI may unlock many benefits for the world, but on its current path, it is also likely to exacerbate existing harms of digital technology, such as the mental health crisis, the erosion of societal trust, and political polarization. Part of the governance challenge is that at the micro level these ills are subtle and widely dispersed; an algorithmically amplified piece of political content is not acutely harmful to an individual user, for instance. It is only at the macro, societal level that the scope and consequences of these local interactions come into view. Virality, in this sense, becomes an enemy of reason.[14]

---

[12] See, W. Naudé, "Artificial intelligence: neither Utopian nor apocalyptic impacts soon," *Economics of Innovation and New Technology* 30:1 (2021): 1-23

[13] https://www.sciencedirect.com/topics/computer-science/preferential-attachment.

[14] https://www.wired.com/story/the-one-internet-hack-that-could-save-everything-section-230/

III. Recommendations

The following recommendations are for national governments, international organizations, researchers, and the participants in the multilateral processes that are underway to facilitate global cooperation for AI safety and equality. Ultimately, the goal of global governance of AI should be to provide a **fit, adaptive response system** that can prevent the perpetuation – or "lock in" – of harmful tendencies and permit the positive aspects of AI to develop and flourish. Complexity points to some general precepts for such adaptive governance.[15]

- Policy makers should identify and understand **feedback loops**. Positive feedback loops increase the effect of a change, while negative feedback loops in general dampen or buffer the effect of a change and help maintain stability in the system (although negative feedbacks too strong can cause overshoot and instability).
- Technology is embedded in larger social, economic, and political systems; governance of it must consider **second-order effects,** the unintended or follow-on consequences that result from the use of technology. All policymakers are trained to look for unintended consequences, but the speed and power of AI requires special precautions.
- Complex adaptive systems have **high-leverage intervention points** at which the system may be able to be flipped from one basin of attraction to another. The architects of the Bretton Woods system, for example, remade a series of key bilateral relationships into a global economic order by creating a most-favored-nation principle.
- The features and behaviors of a system emerge from interactions at **multiple scales** – individuals, collectives, and the macroscopic; governing the system depends on understanding and managing conflicts that emerge as the system expands. A **layered governance model** could work across technical, ethical, social, and global levels with tailored approaches to each layer.[16]

With these precepts in mind, our specific recommendations are organized into three overarching categories: (1) access and power; (2) international relations and global stability; (3) accountability and liability.

(1) Access and Power

How can we ensure that AI benefits the poor as well as the rich, does not worsen inequality, and serves public, non-market-driven interests beyond corporate profit and shareholder returns?

---

[15] For more background, see also: John H. Holland, *Hidden Order: How Adaptation Builds Complexity*, Basic Books (1996); Simon Levin, *Fragile Dominion: Complexity and the Commons*, Basic Books (2000); Anne-Marie Slaughter, *The Chessboard and the Web: Strategies of Connection for a Networked World*, Yale University Press (2017); Melanie Mitchell, *Complexity: A Guided Tour*, Oxford University Press (2011); and Geoffrey West, *Scale: The Universal Laws of Growth, Innovation, Sustainability, and the Pace of Life in Organisms, Cities, Economies, and Companies*, Penguin Press (2017).

[16] See U. Gasser & V. Almeida, "A layered model for AI governance," *IEEE Internet Computing* 21(6) (2017): 58–62.

Embed AI responsibility into the global human rights architecture. The adoption of the UN Declaration on Human Rights more than 75 years ago created a normative cascade, thanks to which basic human rights have become a core consideration of many governance processes. Today, the human rights discourse catalyzed by the Declaration continues to shape fundamental issues around privacy, political space, education, equality, and development. Principles related to responsible, rights-respecting AI should be integrated into the human rights discourse, whether through a global declaration or by other, more creative means. Doing so could help to generate healthy, stabilizing negative feedback loops when risks emerge (e.g. if AI is impinging on rights to privacy or creating large inequalities), gently coaxing the system back into its basin of attraction, and could help to tilt AI toward more beneficial and less destabilizing outcomes for all.

Encourage the creation of national and regional compute resource centers and training data repositories for use by researchers and start-ups. These resources (training data and compute) are high-leverage points in the system. They are the essential resources for AI development. Right now, they are overwhelmingly controlled by a handful of large firms located in the rich world. If nations were to join together to invest in joint centers and repositories, then it could help smaller businesses and nations compete and reduce the concentration of market power in big tech firms that have the resources to develop powerful AI models.

Support the establishment of a robust, global network of public-interest AI research hubs that together can create a public AI infrastructure. Complexity thinking suggests that the greatest risks from AI could arise from the concentration of research and development in the hands of a few. Overly centralized models, or approaches that rely too heavily on a small set of interested actors, could prevent the emergence of healthy negative feedback loops that prevent AI from evolving in ways that produce harmful outcomes. More AI research is needed into specific scientific problems such as renewable energy, basic scientific research, and improved understanding of powerful AI systems themselves.

In the United States, the National Artificial Intelligence Research Resource task force led by the National Science Foundation and the White House Office of Science and Technology Policy has developed a road map for creating "widely accessible AI research cyberinfrastructure that brings together computational resources, data, testbeds, algorithms, software, services, networks, and expertise" that can be used by researchers.[17] Some have argued for the importance of making shared public hardware and software available to AI researchers to study the internal workings and thus improve the explainability of large language models.[18] Though laudable, these efforts

[17] "Strengthening and Democratizing the U.S. Artificial Intelligence Innovation Ecosystem," National Artificial Intelligence Research Resource Task Force, January 2023, https://www.ai.gov/wp-content/uploads/2023/01/NAIRR-TF-Final-Report-2023.pdf

[18] See Benjamin Bucknail and Robert Trager, "Structured access for third-party research on frontier AI models: Investigating researchers' model access requirements," Oxford Martin AI Governance Initiative White Paper, October 2023,

are relatively modest. An international effort on the scale of CERN for particle physics is needed to increase the scale and capital pools for public AI research. Funding might come from globally coordinated wealth taxes, fees levied on excess profits of very large AI companies, or micro-taxes levied on some kinds of AI transactions.[19]

(2) International Relations and Global Stability

What kinds of strategies and global governance frameworks and bodies can help ensure global stability amid disruptions caused by the development and deployment of AI technologies?

AI governance institutions need to be fit for the challenges and nature of AI. While there are many potential models for governing AI, they should be evaluated on the basis of how well they fit AI's specific challenges. These include:

1. non-linear change, as AI jumps in capacities;
2. uncontrollable outcomes, as AI moves across actors quickly and invisibly;
3. proliferation across geographies and domains;
4. concentration of power and resources among a small number of actors; and
5. acceleration, as AI generates increasingly rapid growth in a wide variety of areas and on multiple scales.

A "fit" governance response would therefore need to involve:

1. elasticity, the ability to scale up and out quickly in response to surprising developments;
2. the capacity to track AI risks at multiple levels;
3. transnational authority and capacity that does not become impeded by national boundaries or questions of sovereign control;
4. safeguards against monopolistic tendencies or rapid concentrations of power; and
5. a scientific analytic capacity that can accelerate alongside AI.

One promising proposal that could meet these characteristics would be some version of a Global AI Observatory proposed by Wendell Wallach and Anja Kaspersen.[20] The observatory would be a neutral scientific body that would issue assessments about present and future risks, convene global debate, maintain an incident database, and support responsible, safe AI development. It would include expertise from different fields and attempt to provide an integrated, adaptive, and anticipatory picture of AI development and risks at multiple scales and including second-order effects. The observatory could also identify high-leverage intervention points and thresholds –

---

https://www.oxfordmartin.ox.ac.uk/downloads/academic/Investigating_Researchers%E2%80%99_Model_Access_Oct23-compressed_3.pdf; and Stephen Casper, Carson Ezell, et al., "Black-Box Access is Insufficient for Rigorous AI Audits," arXiv, January 2024, https://arxiv.org/pdf/2401.14446.pdf.

[19] https://www.schneier.com/blog/archives/2023/07/the-ai-dividend.html; https://www.ft.com/content/242c8f5a-43af-43d5-875f-261a0841045a

[20] https://www.carnegiecouncil.org/media/article/the-case-for-a-global-ai-observatory-gaio-2023

the events and metrics that would indicate when the system had entered a state of unacceptable risk. It would need not be a single institution but could be composed of a distributed network of offices located in different regions with connections to different sectors and scales (local, national, regional).

Create, bolster, and maintain international channels and forums for information sharing and transparency on AI risks of mutual concern. As AI develops and plays a greater role in high-risk domains such as military command and control, global stability will depend in part on the existence of channels for feedback and information flow among nations, especially adversarial ones. Whether it's ants touching antennae or financial markets responding to changing prices, all complex systems depend on regular communications among actors. Among great powers, transparent, reliable information flows are necessary to both reduce risk, such as from AI-driven nuclear threats, and advance shared global benefits, such as the proliferation of AI-driven green technologies.

As examples, policymakers could create both bilateral hotlines – so-called "red phones" that are always active – and use international forums such as the G20, the UK AI Safety Summit, or the Pugwash Conferences on Science and World Affairs, in which rivals feel comfortable deliberating and agreeing. In foreign AI policy, democratic nations should think in terms of a "dual-track strategy" that strengthens values-based AI cooperation among open societies, while at the same time building such forums, channels, and norms for cooperation in areas of mutual concern among all nations.[21]

(3) Accountability and Liability

How can the world begin to hold AI developers accountable for the products they create and develop a liability regime around AI, even at the level of norms and standards?

Develop multi-layered AI accountability based on transparency, incentives, and regulations. Accountability can be critical for managing local interactions and conflicts that emerge among scales, particularly between individuals and collectives (such as companies). Liability law that ensures product safety, for instance, can be thought of as a healthy negative feedback loop that prevents a system from producing dangerous outcomes. No such equivalent for AI applications yet exists. If a model's behavior harms an individual or society, the company that created that model is not held liable.

Given the nascent and fast-changing nature of AI systems and products, not to mention political barriers to the regulation of technology companies, policymakers should think of accountability

[21] Allison Stanger et al, "Terra Incognita: The Governance of Artificial Intelligence in Global Perspective," Annual Review of Political Science, 2024.

for AI as a system composed of multiple layers stacked atop one another.[22] One possible model of this system would have three layers.

- The foundational layer would be transparency, which should include plain language disclosure requirements for AI products and improved data collection, dissemination, and actuarial analysis about the impacts of AI systems.
- The next layer would be "soft" regulations, such as voluntary standards and public pressure and other measures that can affect the incentives of AI developers, such as avoiding reputational damage.
- The top layer would be the legal and judicial processes that create and enforce binding liability laws.

Policymakers should focus on strengthening all the layers of the accountability system. And they should avoid passing measures that weaken layers. One measure that should absolutely be avoided is some form of a blanket liability shield for AI systems. Liability is complicated, and no set of regulations can clearly define all the ways that the creators and users of AI systems might be liable. But the judicial process is a robust feedback system in society that can feel out boundaries as a technology progresses. Shielding AI from that process would be a mistake.

*

IN CONCLUSION, ensuring that AI develops safely and beneficially is a hard but achievable task. National and global governance systems will have to be designed in ways that enable them to adapt and learn as the technology develops. The insights that flow from the study of complex adaptive systems are valuable in thinking through the new types of approaches and institutions that will be necessary.

In a sense, AI is the next phase of the digital revolution, which has been developing for decades but has suddenly take a dramatic leap forward. Governments, however, are still struggling to regulate the digital technology of the last several decades. We can learn from our failures to date, but we will need new tools and methodologies as well. The question in the balance is whether human intelligence can use and govern artificial intelligence to avoid the worst dangers and achieve the greatest good for the greatest number. We believe that it can, but only if we can cast off centuries of rules and practice that are slow and hard to adapt. We must learn from the complexity and adaptability of nature itself.

a

[22] One analogue might be the accountability system that has emerged for environmental protection. See, for instance, C. Folke, "Social-ecological systems and adaptive governance of the commons," *Ecological Research* 22 (2007).