

Point of View

Building Blocks for AI Governance



Ian Bremmer and Mustafa Suleyman

Policymakers must adhere to five guiding principles to govern AI effectively



Artificial intelligence will open people's lives and societies to groundbreaking scientific advances, unprecedented access to technology, toxic misinformation that disrupts democracies, and economic upheaval. In the process, AI will trigger a fundamental shift in the structure and balance of global power.

This creates an unparalleled challenge for political institutions around the world. They will have to establish new norms for a dynamic novel technology, mitigate its potential risks, and balance disparate geopolitical actors' interests. Increasingly, these actors will come from the private sector. And it will require a high level of coordination from governments, including strategic competitors and adversaries.

In 2023, governments around the world woke up to this challenge. From Brussels to Beijing to Bangkok, lawmakers are busy crafting regulatory frameworks to govern AI, even as the technology itself advances exponentially. In Japan, Group of Seven leaders launched the "Hiroshima Process" to tackle some of the trickiest questions raised by

AI may become the first technology with the means to improve on itself.

generative AI, while the UN launched a new AI high-level advisory body. At the Group of Twenty summit in New Delhi, Indian Prime Minister Narendra Modi called for a new framework for responsible human-centric AI governance, and European Commission President Ursula von der Leyen advocated for a new AI risk monitoring body modeled on the Intergovernmental Panel on Climate Change.

In November, the UK government hosted the world's first leader-level summit dedicated to addressing AI safety risks. Even in the US, home of the biggest AI companies and traditionally hesitant to regulate new technology, AI regulation is a question of when, not if, and a rare instance of bipartisan consensus.

This flurry of activity is encouraging. In a remarkably short amount of time, world leaders have prioritized the need for AI governance. But agreeing on the need for regulation is table stakes. Determining what *kind* of regulation is just as important. AI doesn't resemble any previous challenge, and its unique characteristics, coupled with the geopolitical and economic incentives of the principal actors, call for creativity in governance regimes.

AI governance is not just one problem. When it comes to climate change, there may be many routes to achieving the ultimate objective of lowering greenhouse gas emissions, but there is a single overriding objective. AI is different, as an

AI policy agenda must simultaneously stimulate innovation to solve intractable challenges and avoid dangerous proliferation, and it must help attain geopolitical advantage without sleepwalking the world into a new arms race.

The AI power paradox

The nature of the technology itself is a further complication. AI can't be governed like any previous technology because it's unlike any previous technology. It doesn't just pose policy challenges; its unique characteristics make solving those challenges progressively harder. That is the AI power paradox.

For starters, all technologies evolve, but AI is hyper-evolutionary. AI's rate of improvement will far surpass the already powerful Moore's Law, which has successfully predicted the doubling of computing power every two years. Instead of doubling every two years, the amount of computation used to train the most powerful AI models has increased by a factor of 10 every year for the past 10 years. Processing that once took weeks now happens in seconds. The foundation technologies that enable AI are only going to get smaller, cheaper, and more accessible.

But AI's uniqueness is not just about expanded computing capacity. Few predicted AI's evolution, from its ability to train large language models to being able to solve complex problems or even compose music. These systems may soon be capable of quasi-autonomy. This would on its own be revolutionary but would come with an even more dramatic implication: AI may become the first technology with the means to improve on itself.

AI proliferates easily. As with any software, AI algorithms are far easier and cheaper to copy and share (or steal) than physical assets. And as AI algorithms get more powerful—and computing gets cheaper—such models will soon run on smartphones. No technology this powerful has ever been so accessible so widely so quickly. And because its marginal cost—not to mention marginal cost of delivery—is zero, once released, AI models can and will be everywhere. Most will be safe; many

have been trained responsibly. But, as with a virus, all it takes is one malign or “breakout” model to wreak havoc.

Incentives point toward ungoverned AI

The nature of AI suggests different incentives as well. Dual-use technologies are nothing new (there's a reason civilian nuclear proliferation is closely monitored), and AI is not the first technology whose civil and military uses are blurred. But whereas technologies such as nuclear enrichment are highly complex and capital-intensive, AI's lost cost means it can be deployed endlessly, whether for civil or military use. This makes AI more than just software development as usual; it is an entirely new and dangerous means of projecting power.

Constraining AI is hard enough on a technological basis. But its potential for enriching and empowering powerful actors means that governments and the private companies developing AI are incentivized to do the opposite. Simply put, AI supremacy is a strategic objective of every government and company with the resources to compete. If the Cold War was punctuated by the nuclear arms race, today's geopolitical contest will likewise reflect a global competition over AI. Both the US and China see AI supremacy as a strategic objective that must be achieved—and denied to the other. This zero-sum dynamic means that Beijing and Washington focus on accelerating AI development, rather than slowing it down.

But as hard as nuclear monitoring and verification were 30 years ago, doing the same for AI will be even more challenging. Even if the world's powers were inclined to contain AI, there's no guarantee they'd be able to, because, as in most of the digital world, every aspect of AI is currently controlled by the private sector. And while the handful of large tech firms that currently control AI may retain their advantage for the foreseeable future, it is just as likely that the gradual proliferation of AI will bring more and more small players into the space, making governance more complicated. Either way, the private businesses and individual technologists

who will control AI have little incentive to self-regulate.

Any one of these features would strain traditional governance models; all of them together render these models inadequate and make the challenge of governing AI unlike anything governments have faced before.

Governance principles

If global AI governance is to succeed, it must reflect AI's unique features. And first among those is the reality that as a hyper-evolutionary technology, AI's progress is inherently unpredictable. Policymakers must consider that given such unpredictability, any rules they pass today may not be effective or even relevant in a few months, let alone a few years. To box in regulators with inflexible regimes now would be a mistake.

Instead, good governance would be best served by establishing a set of first principles on which AI policymaking can be based:

- **Precautionary:** The risk-reward profile of AI is asymmetric; although there are vast benefits to AI's potential, policymakers must guard against its potentially catastrophic downsides. The already widely used precautionary principle needs to be adapted to AI and enshrined in any governance regime.
- **Agile:** Policymaking structures tend to be static, prizing stability and predictability over dynamism and flexibility. That won't work with a technology as unique as AI. AI governance must be as agile, adaptive, and self-correcting as AI is fast-moving, hyper-evolutionary, and self-improving.
- **Inclusive:** The best industry regulation, especially when it comes to technology, has always worked collaboratively with the commercial sector, and this is especially true for AI. Given the exclusive nature (at least for now) of AI development—and the complexity of the technology—the only way for regulators to properly oversee AI is to collaborate with private technology companies. To reflect the borderless nature of AI,

governments should make companies parties to international agreements. Including private companies in high diplomacy may veer toward unprecedented, but excluding those who have so much control would doom any governance structure that excludes them before it even starts.

- **Impermeable:** For AI governance to work, it must be impermeable; given AI's ability to easily proliferate, just one defection from the regime could allow a dangerous model to escape. Therefore, any compliance mechanisms should be watertight, with easy entry to compel participation and costly exit to deter noncompliance.
- **Targeted:** Given AI's general-purpose nature and the complexities involved in governing it, a single governance regime is insufficient to address the various sources of AI risk. In practice, determining which tools are appropriate to target which risks will require developing a live, working taxonomy of discrete potential AI impacts. AI governance must therefore be targeted, risk-based, and modular rather than one-size-fits-all.

Governing AI will be among the international community's most difficult challenges in the coming decades. As important as the imperative to regulate AI is the imperative to regulate it correctly. Current debates on AI policy too often tend toward a false debate between progress and doom (or geopolitical and economic advantages versus risk mitigation). And rather than think creatively, solutions too often resemble paradigms for yesterday's problems. This will not work in the age of AI.

Good policymaking will be vital, but getting there rests on good institutions. To build these institutions, the international community will need to agree on a conceptual framework for how to think about AI. We offer these principles as a start. **F&D**

IAN BREMMER is president and founder of Eurasia Group and GZERO Media. **MUSTAFA SULEYMAN** is CEO and cofounder of Inflection AI.

Fostering More Inclusive Democracy with AI



Hélène Landemore

AI can enhance democratic institutions by ensuring citizens' voices are truly heard

People worry that artificial intelligence is, or will soon be, undermining democracy. They fear AI will take away jobs, destabilize the economy, and widen the divide between the rich and the poor. This could further concentrate power in the hands of a few tech companies and weaken government structures designed to regulate them. Some also fear that tech giants and government may increasingly delegate human decision-making to machines, eventually replacing democracy with “algocracy,” rule not by the people but by algorithm.

This dystopian vision misses our current capacity to shape AI development. We, as human societies, have the political ability (at least for now) and the responsibility to address the harm AI could inflict on us. We also have the technological opportunity to harness AI to enhance our democracy in a way that strengthens our collective ability to govern—rather than simply regulate—AI.

Like other ethical and political challenges, such as gene editing, AI governance requires not just more expert intervention and regulation but more citizen voice and input—for example, on how to navigate the distributive impact of AI on the economy. Like other global concerns, such as climate change, AI governance requires this democratic voice to be heard at the level of international institutions. Luckily, AI has the potential to usher in a more inclusive, participatory, and deliberative form of democracy, including at the global scale.

Participatory experiments

For 40 years many governments have engaged in experiments aiming to include ordinary citizens in policymaking and lawmaking in richer ways than through voting alone. These experiments have mostly been local and small-scale, much like the citizens' assemblies and juries that have proliferated on climate and other issues. A 2020