

# **SPECIAL** REPORT no. 229

## **‘Moving Horizons’: A Responsive and Risk-Based Regulatory Framework for A.I.**

**Samir Saran, Anulekha Nandi, and Sameer Patil**



**JUNE 2024**

# Abstract

**A**s Artificial Intelligence (AI) capabilities constantly evolve, its regulation can no longer remain simply an exercise in optimisation and mitigation—or maximising innovation opportunities and minimising the risk of harm. AI’s intersecting socio-economic and legal implications require dynamic governance arrangements to identify, respond to, and anticipate continually shifting regulatory imperatives. This report makes a case for a framework that not only anticipates, recognises,

and assesses risks, but also responsively manages them.<sup>a</sup> Such a framework would open up pathways for harmonising sovereign imperatives of building national competencies while fostering multilateral cooperation on developing globally accepted standards to facilitate the responsible deployment of AI innovation.

Attribution: Samir Saran, Anulekha Nandi, and Sameer Patil, “Moving Horizons’: A Responsive and Risk-Based Regulatory Framework for A.I.,” *ORF Special Report No. 229*, June 2024, Observer Research Foundation.

---

a This report is based on the keynote address delivered by Dr Samir Saran in April 2024 during the event, ‘AI Policy Symposium: A Global Stocktaking’, hosted by the Stanford Institute for Human-Centered AI, Stanford University, California.

# Introduction

**A**rtificial Intelligence (AI) systems have come a long way since 2016, when Microsoft released Tay, an AI chatbot which had to be shut down within a day after spewing racist and anti-Semitic tweets.<sup>1</sup> In 2022, OpenAI introduced ChatGPT, beginning a new era for generative AI where algorithms could churn out diverse content at scale; its champions say it could contribute trillions of dollars to the world economy.<sup>2</sup> The latest version of ChatGPT, at the time of writing, can process and produce information across different modalities like text, image, and video.<sup>3</sup>

If the Tay experience taught the world anything, however, it is that there is a need for guardrails for AI algorithms that learn dynamically and interactively from harmful user behaviour or

that draw patterns and inferences from widely prevalent human conduct with negative social, economic and political consequences. While AI's potential for generating explicitly harmful outputs may have reduced since Tay, its effects have become less apparent as a result of the widespread use and re-use of common datasets with inherent biases across different algorithms and models.<sup>4</sup> Such consequential but less visible harms have come to be entrenched and pervasive as companies continue to develop and embed AI capabilities in their products, services, processes, and decision-making.<sup>5</sup>

For instance, concerns have amplified about pervasive gender and racial bias and discrimination in AI algorithms.<sup>6</sup> Some AI-enabled facial recognition systems in the United States (US) have underperformed when presented with darker skin tones<sup>7</sup> and criminalised historically marginalised groups.<sup>8</sup> Similarly, studies in the US provide insights on how natural language models are perpetuating stereotypes particularly for identities at the intersection of gender, ethnicity, and race:<sup>9</sup> hiring algorithms are discriminating against protected ones spanning religion, race, gender, and disability,<sup>10</sup> and credit assessment algorithms are marginalising women.<sup>11</sup> These emerging risks are exacerbated by a developer pool that is not diverse enough to consider the experiences of under-represented groups who are missing from consequential decision-making within the AI pipeline.<sup>12</sup> This is compounded by the fact that countries, trying to develop national competencies in AI, have to grapple with systemic issues of data consolidation and concentration of computing infrastructure within large transnational tech companies based in the US.<sup>13</sup>

Concurrently, as AI systems acquire increasingly autonomous capabilities, it raises questions around ownership and authorship of intellectual property (IP). Beginning in 2018, US computer scientist Stephen Thaler filed a series of IP applications across jurisdictions to

designate his AI system DABUS (Device for the Autonomous Bootstrapping of Unified Sentience) as an inventor. His applications, sometimes more than one in a single jurisdiction, were rejected by Australia, the United Kingdom, the US, New Zealand, as well as the European Union Patent Office, which all argued that authorship of an invention can only be vested in a legal person. South Africa was the only jurisdiction that granted him favour, when in July 2021, its patent office deemed DABUS an “inventor”.<sup>14</sup>

In India, the Indian Copyright Office in 2020 rejected an application of an AI system RAGHAV to be regarded as a co-author of an artistic work.<sup>15</sup> Subsequently, another application where RAGHAV was listed as co-author along with its human creator was accepted.<sup>16</sup> However, the office later issued a withdrawal notice asking the human co-author to inform the office about the legal status of the AI tool.<sup>17</sup>

Such conundrums around IP ownership bring the attention back to the location of responsibility and liability of AI-generated outputs, more importantly in cases of adverse consequences. The very nature of AI systems—opaque, inscrutable, and autonomous—presents massive challenges



for determining liability rules. How does one trace causality and assign fault when producers are unable to foresee them due to the self-learning nature of algorithms?<sup>18</sup> The fluidity of these algorithms defy traditional definitions of “product defects”, which in turn complicates the determination of liability within a given supply chain. Other approaches, for instance, defining a principal–agent relationship, would ascribe liability to the deployer and therefore have limited uptake.<sup>19</sup> Meanwhile, more radical approaches of granting AI legal personhood would have to grapple with ancillary questions including whether AI can own property or enter into contracts in its name.<sup>20</sup>

These instances raise more questions than answers: How would liabilities be established for AI-driven harms that stem from the perpetuation of systemic bias and suffer from attribution of blame and apportionment of liabilities? Should it be done be through a product liability regime, principal-agent relationship, or newer legal contours of AI-society interface? How should we deal with systemic issues of under-representative datasets that result from the historical over-representation of men in education and employment,<sup>21</sup> underrepresentation of women in drug trials, or the failure of automobile safety

tests to account for the dimensions of women’s bodies?<sup>22</sup> How do we engage with emerging risks of algorithmic hallucinations, dynamic learning of discriminatory behaviour that might be a combination of systemic issues within its data and myriad user interactions? How should we *ex post* address the market concentration enjoyed by data-rich Big Tech firms enjoying the cascading benefits of being first movers of Web 2.0?

Indeed, AI governance and regulation is like the many-headed Hydra, presenting policymakers and regulators with persistent and seemingly intractable challenges. The development of AI involves an entire ecosystem of stakeholders and conditions with differential control and distribution of resources.<sup>23</sup> AI also defies the clear delineation of cause and effect or direct evaluation of potential harms given its opaque nature, general-purpose application, and cross-effects across domains. These leave regulators grappling with thorny issues to determine where individual rights end, proprietary rights begin, and AI rights take over.

Regulation thus ceases to be a matter of optimisation–mitigation, with AI having cascading socio-economic and legal implications. An attempt

to regulate certain aspects of AI may succeed, but goalposts shift and solutions to respond to them have to adapt in sync with the evolution of AI.

This report offers the ‘Moving Horizons’ framework as a plausible guide to AI regulation. Drawing on elements of responsive and risk-based regulation, ‘Moving Horizons’ aims to capture evolving and emerging regulatory trends and incorporate elements of existing regulations that in themselves may become outdated. This approach recognises the underlying conditions and shifts in

the landscape and uses them to balance, on one hand, the opportunities for innovation, and on the other, the risks of adverse consequences. This involves a dynamic governance outlook, building institutional capabilities, streamlining processes through adaptive regulatory approaches, and developing regulatory and technical competencies that allow regulators to respond to changing environmental and ecosystem needs.<sup>24</sup>

# Underlying Conditions and Shifts

**I**t is important to outline the most crucial underlying conditions and important shifts that frame governance priorities around AI. These can be arranged and discussed in the following three broad categories.

## **The tragedy of data and digital commons**

Data is not like the traditional commons: it is not a finite resource nor is it naturally endowed, and is non-rivalrous.<sup>25</sup> However, proprietary algorithms and computing infrastructure used to scrape and harvest the web for publicly available data subvert the nature of personal data ownership.<sup>26</sup> This serves to undermine existing IP regimes through non-personalisation of personal data at the rate and scale at which Big Tech operates.<sup>27</sup>

This stands to be amplified through existing repositories of user-generated content. With data collected and collated in composite datasets and imbued with algorithmic analysis, the original data

is transformed into an analytical output over which the individual data owner ceases to have any right. This has diluted original ownership and accrual of benefits to users, with no remedial measures or redistribution of the extracted economic value to account for negative externalities and adverse consequences.<sup>28</sup>

## **Generative dilemmas and the global AI economy**

AI development depends on large volumes of data and computational capacity. The data troves possessed by Big Tech companies like Microsoft, Facebook, Amazon, and Google have enabled them to train large AI models with positive feedback loops.<sup>29</sup> This has raised anti-competitive concerns in many jurisdictions, predominantly the European Union (EU).<sup>30</sup> These systemic issues give rise to newer predicaments that need to be managed—i.e., the dilemma of competition or collaboration, and the current market structure within AI development and deployment that enables or constrains that.

At present, there is limited work on incorporating consumer-facing evaluation metrics for transparency and accountability to inform consumers of the AI system's risks in the form of bias and discrimination that can be encountered upon use.<sup>31</sup> Given that the incorporation of evaluation algorithms increases computational and development cost, companies are faced with the choice of either securing their systems to build trustworthy AI or maximising the speed of innovation without guardrails. The dilution of national and plurilateral capabilities to legislate and the demand for a multistakeholder approach on tech governance, broadly, and AI specifically, have created largely unregulated and ungoverned spaces that do not provide regulatory signals for good behaviour by corporations. While leaders of Big Tech firms have, at least in principle, endorsed the idea of regulating AI, there is little consensus on what such regulation should look like.<sup>32</sup> Therefore, more attention is needed on the speed of deployment, the regulatory interface (or the lack of it), and the costs and consequences of such rapid deployment.<sup>33</sup>

### **Skin in the AI game**

Globally, the US and China are the current hubs of AI innovation; the EU, meanwhile, having adopted its landmark AI Act in 2024, stands at the forefront of AI regulation. This distance between the centres of innovation and regulation foreshadows the extra-territorial scope with which EU AI regulations operate.<sup>34</sup> Particularly, the developing world is largely relegated to being data providers to the more advanced centres of innovation where much of the revenues and economic benefits accrue.<sup>35</sup> This leads to conditions where the developing world ends up providing either critical resource for AI innovation or becoming the subject of AI regulation which constrains its competitive capacity to innovate and participate in the global AI economy.

AI development thus continues to be both extractive and skewed, and those that contribute to the development and evolution of AI solutions with their personal data remain unserved or marginally served. Consequently, given these nodes of influence, corporations and platforms engage with key capitals such as Brussels in the developed world on regulation while ignoring the rest, further deepening the gaps in capabilities and capacities.

# Market Perversions

**T**he conditions currently undergirding regulatory concerns on AI are reinforced by perversions in the market. This section outlines two case studies—the defence sector and the sex robot industry—to illustrate the negative externalities engendered by AI systems and the deep and pervasive inroads they have made so far.

## **Defence sector**

Global demand for dual-use emerging technology-driven military capabilities is witnessing a surge and is increasingly being fulfilled by firms outside the defence-industrial base.<sup>36</sup> In the US, for instance, defence tech start-ups have proliferated over three phases in the last two decades, with the most recent wave in 2018-2023 dominated by emerging technologies like AI/Machine Learning.<sup>37</sup> Major powers are making investments in developing military applications of AI.

For instance, the US Department of Defense has made a US\$1.8-billion budget request in 2024 for AI, with China spending just a tad less, at US\$1.6 billion yearly on AI.<sup>38</sup> Increased investment in AI start-ups by defence players has caused the market to reach a value of US\$9.23 billion in 2023.<sup>39</sup> The autonomous counter-drone systems are poised to be operationally ready in 2024, with the market expected to reach US\$2.1 billion.<sup>40</sup> Devising autonomous and AI-powered smart munitions, systems, and weapons will shape both the design and principles on which AI may be regulated or made accountable.

## **Secretarial and sex robot sector**

The most prolific consumer-facing application of robots so far has been in the form of voice assistants and sex companions.<sup>41</sup> While Amazon's Alexa has been programmed to refrain from engaging with questions of sexually explicit or harassing nature,

developers have reported that at least 5 percent of user interactions with the voice assistant are unambiguously sexually explicit.<sup>42</sup> That voice assistants are gendered as female, assigned secretarial roles, and programmed to give docile responses to verbal abuses highlight the deep biases in real life that seep into the virtual.<sup>43</sup>

In 2022, the sex robot industry was valued at US\$200 million, with an estimated 56,000 sex robots sold per year, each priced anywhere between US\$5,000 and 15,000.<sup>44</sup> Psychosexual therapists say sex robots can be beneficial for those who find intimate relationships difficult or are surviving trauma.<sup>45</sup> At the same time, researchers

warn that these robots increase the objectification of women and children and alter perceptions of consent that would otherwise be considered illegal, with some models simulating rape scenarios.<sup>46</sup>

This raises pertinent questions about differential regulation of civil and military uses for dual-use technologies, and implications for civil and criminal law. These concerns are undercut by two other driving factors as discussed in the succeeding section.



# Consolidation and Liabilities

**T**he consolidation and centralisation of innovation in AI comes on the heels of the transformations that marked the advent and evolution of the Web 2.0. During this time, social media companies became repositories of large volumes of user-generated content which are reinforced through the continued use of such platforms. These data troves become the foundational basis on top of which such platforms are able to scale and drive demand for their cloud computing infrastructure services.<sup>47</sup>

AI arguably marks the end of decentralised innovation of the internet. Earlier, defence research such as in the case of the United States' Defense Advanced Research Projects Agency was outsourced to a number of different

companies. The resulting architecture of the internet facilitated multi-nodal innovation with decentralised computational advancements powering the advent of digital societies.<sup>48</sup> AI innovation is, however, marked by centralisation, given large data pools and computing capabilities needed for the development of AI models. Both of these are possessed by a few, predominantly the Big Tech firms that dominated the previous rounds of digitalisation. This enabled them to acquire and capitalise on massive data troves of user-generated content. Thus, they have the capacity and resources to invest and build overwhelming compute capabilities even as they mine vast reservoirs of data. This presents newer entrants with an unequal and a near insurmountable playing field. It is a vicious cycle as all newcomers need to depend on compute capabilities offered by the large corporations and rely on their infrastructure, such as cloud services,

which drives further consolidation.<sup>49</sup> In sum: the large capital investments required for developing foundational models and computational capacity inhibits market entry in this space.<sup>50</sup> Yet, green shoots are slowly appearing through local ecosystem investment, open-source models, and institutional support and funding.<sup>51</sup> Are these sufficient and rapid enough? And how will this current business environment affect the important questions around accountability?

This aspect is important. Given the risks stemming from AI systems and lack of consensus on an established liabilities regime, companies could shield themselves behind an invisible safe harbour of their own making wherein they could dismiss any attribution of risks of harms to themselves and avoid acknowledging any responsibility. Intermediary liability protections,

in a new avatar, may absolve AI providers of harmful product development and service provision. According to Google's generative AI additional terms of service, for example, the company will not be claiming ownership of content generated by AI systems.<sup>52</sup> While the aim of the policy is to allow users to claim ownership without any copyright issues, it complicates designation of liabilities in adverse consequences. In general, the legal architecture is unable to keep pace with the rapid developments in the AI space and there is a real fraying of both domestic and international regimes. Interested parties are perhaps investing in this state of flux with respect to the legal approaches for governing AI.

# Ethics as Obfuscation of International Law

**A**s a case in point, global AI governance and regulations have tended to be limited to AI ethics as a mode; these ethical principles bypass systems of international law by virtue of their being non-binding.<sup>53</sup> Consequently, such toothless principles either become meaningless as they work at cross-purposes with technical realities (e.g., the dichotomy of preserving privacy while ensuring representative datasets), remain isolated with narrow sectoral focus, or lack consequences because of their non-binding nature. The translation of normative ethical prescriptions into technical codification poses challenges. For instance, the treaty on AI adopted by the Council of Europe in May 2024—the first legally binding treaty of its kind—lacks clear specifications for the delineation of obligations beyond adherence to normative

principles.<sup>54</sup> The treaty does not explicitly address questions around liabilities and responsibilities, which are important for redressing harms arising out of AI systems.<sup>55</sup>

For its part, the United Nations General Assembly adopted a resolution on AI early this year,<sup>56</sup> but stopped short of proposing some form of discussion on changes that may be needed to international law. This is particularly important given the transnational nature of the AI systems, the predominance of the English language as a form of data on which AI systems are trained, and under-representation within the AI development life cycle along with global inequalities in national AI competencies.

# The ‘Moving Horizons’ Framework for Responsive and Risk-Based Regulation

**T**he regulation of AI development has to contend with multiple, interrelated realities as discussed in the earlier sections of this report. These require the management of systemic conditions that pervade the system in the form of market concentration, unequal distribution of resources, or under-representation in datasets and the developer community. Systemic conditions originate from multiple sources, affect different actors within the ecosystem, and propagate rapidly, resulting in a domino effect.<sup>57</sup> This highlights the nature of pervasive harms within AI systems, as a result of which it becomes difficult to determine the exact source of emerging risks as they get enmeshed in algorithms, models, and self-learning AI systems. These conditions within its design process shape how and on what AI models are trained that lead to the dynamic emergence of

risks as algorithms continue to learn through user interactions. These include AI hallucinations, and biased, discriminatory, or toxic outputs.

While the EU AI Act utilises a risk classification mechanism, risk management in practice becomes a matter of attention shaping and intervention.<sup>58</sup> Moreover, risk-based regulatory approaches, while being implemented, are affected by differing levels for risk tolerance across jurisdictions and sectors and are shaped by different interpretations of normative principles for evaluation.<sup>59</sup> However, technological and AI systems can contain residual risks even when significant operational controls are being implemented.<sup>60</sup> This highlights the need to nurture dynamic governance capabilities to understand and respond to the convergence and intersection of systemic conditions and emerging

risks with stakeholder, sectoral, societal, and state choices. This is also important for balancing the competing concerns of fostering innovation in tandem with risk mitigation, thereby reinforcing the need for strategic alignment of resources, conditions, and actors.

AI development tends to proceed through an ecosystem of stakeholders with the triple helix of government, academia, and industry.<sup>61</sup> Moreover, systemic conditions and emerging risks highlight the importance of a citizen-centric approach. This can help ensure that regulation can proceed in a responsive and deliberative manner as states develop dynamic capabilities to deal with evolving challenges from AI innovation while taking steps to address systemic conditions.

The ‘Moving Horizons’ regulatory framework draws from responsive regulation’s pyramids of support and sanctions—i.e., addressing adverse consequences while working to expand strengths and the promised potential of AI systems.<sup>62</sup> Given the consideration of managing innovation and risk, it reiterates the importance of developing dynamic regulatory capabilities to identify the level of sanction or support required by actors to bring balance to optimising innovation and minimising risks.<sup>63</sup> In parallel, risk-based regulation involves evaluation of the likelihood and severity of harm. The framework aims to work towards responsive

management of risk and innovation. This helps develop and maintain institutional integrity while taking into account changes in the prevailing landscape.<sup>64</sup>

A ‘Moving Horizons’ regulation approach involves the following components:

**Dynamic governance capabilities and strategic alignment:** Policymakers and regulators need to be in a position to deploy dynamic capabilities to sense, plan, and reconfigure competencies in response to AI innovation. This would involve identifying the problem area to be addressed and the expected incidence of impact, and outlining the regulatory scope. This would ensure it identifies the right stakeholders and mobilises the appropriate government department in visualising response to evolving AI problems and align it to current regulatory demands and social and economic concerns.

**Mapping risks, effects, and responsibilities:** The dynamic delineation of scope would then proceed through the funnel of mapping and classifying risks, severity, causes, and effects. This would then help identify, apportion, and ascribe liabilities and responsibilities to stakeholders involved in the ecosystem.

**Developing frameworks of compliance and support:**

Once risks and responsibilities have been identified, frameworks and standards need to be developed to help businesses demonstrate their efforts to mitigate risks and minimise harm. This would involve developing procedural guidance, compliance frameworks, and standards and benchmarks for reporting through multi-stakeholder consensus between governments, businesses, users, and academia. However, standards and benchmarks need to be iteratively revised based on stakeholder feedback and evolution of technological capabilities.

**Identifying modes and processes of networked escalation:**

The transnational power of big corporations and global inequalities in the distribution of resources for AI production and development highlights the significance of

networked escalation.<sup>65</sup> Signalling to stakeholders the governance capacity to implement escalations engenders more cooperative behaviour towards addressing capacity deficits. Depending on the risk classification and severity, it can begin with self-regulation, then network into non-state regulators like industry and professional bodies if self-regulation fails, followed by networking into established regulatory and government bodies aligned to the issue, to finally moving to shut down or terminate. This requires developing institutional capacity involving both traditional regulatory expertise in conjunction with technical expertise in AI.



# Conclusion: Towards Responsive Harmonisation

**A**s countries and institutions converge on key principles that must govern the development and use of AI, the Organisation for Economic Co-operation and Development (OECD) and the EU principles find resonance in many country position papers.<sup>66</sup> However, as the development of AI is not uniform across nations, there is an AI regulatory divide between the Global North and Global South. As a consequence, there is an over-reliance on regulatory signals from developed world institutions that may not be contextually relevant for all societies. This highlights a need for multilateral initiatives and strategies that upstream sovereign imperatives for national AI competencies in line with global standards, principles, and frameworks.

To begin with, the focus of the developed world is on either AI innovation—like in the US and China—or else regulation, like in the EU. Meanwhile, in countries of the Global South, such as Brazil, Argentina, and India, AI strategies are striving, despite modest budgets, to build national

competencies to drive multi-sectoral innovation. India's national AI strategy, for instance, aims to build responsible AI ecosystems that both foster innovation and drive responsible development through safety and reliability, non-discrimination, privacy and security, and transparency.<sup>67</sup> However, given the disconnect between high-level ethical principles at the global level and focus on developing national competencies at the regional and national level, harmonised and responsive regulation in the form of sovereign aims and international standards become key to sustainable AI governance.

To sum up, the 'Moving Horizons' regulation is an analytical and agile approach that takes responsive and risk-based regulation of AI as a point of departure while recognising the need to manage the convergence of both systemic conditions and emerging risks. To succeed, however, it needs to incorporate the following considerations:


**Regulation of consequences:** In the non-digital world, regulation of risks like vehicular accidents involves both physical and regulatory measures including the installation of speed breakers, imposing speed limits, and restrictions on heavy goods vehicles during certain times of the day. In AI development and deployment, the management of systemic conditions and emerging risks have called into question the heuristic human tendency to control consequences. This is because AI regulation needs to act both as a catalyst for innovation and a deterrent to risk. It reinforces the role of responsibility and obligation to develop, innovate, and manage the convergence of multidimensional nature of risks within AI governance. It proposes a mode of regulation wherein rule-making for responsible innovation is informed and guided by a principled approach rather than normative ideals becoming an abdication of rule-making. This becomes particularly important as developing countries grapple with global inequalities in terms of resources for AI development while trying to protect citizens from risks of harms from proliferating applications in multiple sectors. It highlights the need to establish procedural frameworks and standards of safety evaluation.

**The pandemic paradigm:** During the pandemic, time-critical and life-saving vaccines against COVID-19 went through a three-stage testing process. This included sandboxes to test the innovation in a controlled environment, followed

by population-scale testing, and only then, moving to commercial applications. This was done to prevent unintended and adverse consequences even when lives were at stake. The three-tier ‘innovation to market’ tech absorption framework provides a blueprint for safety and quality control and helps establish procedural guardrails or speed-breakers for profit-maximising innovation at the cost of safety and security.<sup>68</sup> This could include, for example, highlighting the consequences of importing foundational AI models in local and contextual applications and how layers of algorithms on top of it come to determine the nature of its effect for local populations.

**Algorithmic accountability:** Procedural guardrails can only be effective when complemented by suitable frameworks of evaluation. This requires the establishment of standards, benchmarks, and audit mechanisms to institute systems of accountability and transparency which become necessary to designate systems as operationally safe. These require documentation and traceability, like in the case of financial audits with annual or periodic evaluations to detect, manage, and mitigate emerging AI harms. While a number of technical and evaluative algorithmic auditing approaches do exist, these are in dire need of standardisation to establish suitable compliance practices.

A lack of universal standards and frameworks continues to be a crucial challenge underlying inconsistencies in AI governance worldwide. This highlights the importance of forging multilateral cooperation, at one level, to arrive at internationally acceptable and harmonised standards. At the same time, it necessitates national governments to evaluate and develop regulatory mechanisms to build institutional capacities and frameworks

to harness and manage AI-driven transformations currently underway in their jurisdictions. The 'Moving Horizons' approach provides a template for countries seeking to harness AI's potential for positive impact while mitigating potential harms due to the dynamic and emerging risks of AI systems. 

# Endnotes

- 1 Jane Wakefield, "Microsoft Chatbot is Taught to Swear on Twitter," *BBC*, March 24, 2016, <https://www.bbc.com/news/technology-35890188>; Alex Hearn, "Microsoft Scrambles to Limit PR Damage over Abusive AI Bot Tay," *The Guardian*, March 24, 2016, <https://www.theguardian.com/technology/2016/mar/24/microsoft-scrambles-limit-pr-damage-over-abusive-ai-bot-tay>
- 2 PwC, *Sizing the prize: PwC's Global Artificial Intelligence Study: Exploiting the AI Revolution*, PwC Global, 2017, <https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html>.
- 3 OpenAI, "Creating Video from Text," OpenAI, <https://openai.com/index/sora/>
- 4 Fabian Lütz, "Gender Equality and Artificial Intelligence in Europe. Addressing Direct and Indirect Impacts of Algorithms on Gender-Based Discrimination," *ERA Forum* 23, (2022), <https://link.springer.com/article/10.1007/s12027-022-00709-6#Fn35>
- 5 François Cadelon, Rodolphe Charme di Carlo, Midas De Bondt, and Theodoros Evgeniou, "AI Regulation is Coming," *Harvard Business Review*, September – October, 2021, <https://hbr.org/2021/09/ai-regulation-is-coming>
- 6 Ardra Manasi, Subadra Panchanadeswaran, Emily Sours, and Seung Ju Lee, "Mirroring the Bias: Gender and Artificial Intelligence," *Gender and Technology in Development* 26, no. 1 (2022), 1-11; Paula Halls and Debbie Ellis, "A systematic review of socio-technical gender bias in AI algorithms," *Online Information Review* 47, no. 7 (2023).
- 7 Joy Buolamwini and Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification" (paper presented at Proceedings of Machine Learning Research Conference on Fairness, Accountability, and Transparency, 2018)
- 8 Thadeus Johnson and Natasha Johnson, "Police Facial Recognition Technology Can't Tell Black People Apart," *Scientific American*, May 18, 2023, <https://www.scientificamerican.com/article/police-facial-recognition-technology-cant-tell-black-people-apart/>
- 9 Ayanna Howard, "Real Talk: Intersectionality and AI," *MIT Sloan Management Review*, August 24, 2021, <https://sloanreview.mit.edu/article/real-talk-intersectionality-and-ai/>; Wei Guo and Aylin Caliskan, "Detecting Emergent Intersectional Biases: Contextualized Word Embeddings Contain a Distribution of Human-like Biases" (paper presented at 2021 AAAI/ACM Conference on AI, Ethics, and Society, New York, USA, 2021); Tolga Bolukbasi, Kai-Wei Chang, James Zou, Venkatesh Saligrama, and Adam Kalai, "Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings" (part of Advances in Neural Information Processing Systems 29, 2016). [https://proceedings.neurips.cc/paper\\_files/paper/2016/hash/a486cd07e4ac3d270571622f4f316ec5-Abstract.html](https://proceedings.neurips.cc/paper_files/paper/2016/hash/a486cd07e4ac3d270571622f4f316ec5-Abstract.html)
- 10 Miranda Bogen, "All the Ways Hiring Algorithms Can Introduce Bias," *Harvard Business Review*, May 06, 2019, <https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias>; Charlotte Lytton, "AI Hiring Tools May be Filtering Out the Best Job Applicants," *BBC*, February 18, 2024, <https://www.bbc.com/worklife/article/20240214-ai-recruiting-hiring-software-bias-discrimination>; Khari Johnson, "Feds Warn Employers Against Discriminatory Hiring Algorithms," *Wired*, May 16, 2022, <https://www.wired.com/story/ai-hiring-bias-doj-eccc-guidance/><https://www.wired.com/story/ai-hiring-bias-doj-eccc-guidance/>
- 11 "Incident 92: Apple Card's Credit Assessment Algorithm Allegedly Discriminated against Women," *AI Incident Database*, November 11, 2019, <https://incidentdatabase.ai/cite/92#6048603491dfd7f7ac0470be>

- 12 Valentine Goddard, Eleonore Fournier-Tombs, Mercy Atieno Odongo, Jane Ezirigwe, Daniela Chimisso dos Santos, Sarah Moritz, Blair Attard-Frost, and Millicent Ochieng', "Gender Equality and the Environment in Digital Economies" (policy brief prepared for United Nations' 8th Multi-stakeholder Forum on Science, Technology and Innovation for the Sustainable Development Goals, New York, May 2023); Mark West, Rebecca Kraut, and Han Ei Chew, *I'd Blush if I Could: Closing Gender Divides in Digital Skills through Education*, UNESCO and EQUALS Skill Coalition, 2019, <https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1>
- 13 Amba Kak and Sarah Myers West, Eds., *AI Nationalism(s): Global Industrial Policy Approaches to AI*, AI Now Institute, 2024, pp. 8-9, <https://ainowinstitute.org/wp-content/uploads/2024/03/AI-Nationalisms-Global-Industrial-Policy-Approaches-to-AI-March-2024.pdf>
- 14 Rajiv Sharma and Ninad Mittal, "Artificial Intelligence Lacks Personhood To Become The Author Of An Intellectual Property," *LiveLaw.in*, September 22, 2023, <https://www.livelaw.in/law-firms/law-firm-articles-/artificial-intelligence-intellectual-property-indian-copyright-act-singhania-co-llp-238401>
- 15 Aparajita Lath, "AI Art and Indian Copyright Registration," *SpicyIP*, October 10, 2022, <https://spicyip.com/2022/10/ai-art-and-indian-copyright-registration.html>
- 16 Aparajita Lath, "AI Art and Indian Copyright Registration"
- 17 Sukanya Sarkar, "Exclusive: Indian Copyright Office Issues Withdrawal Notice to AI Co-Author," *ManagingIP*, December 13, 2021, <https://www.managingip.com/article/2a5d0jj2zjo7fajsjwwlc/exclusive-indian-copyright-office-issues-withdrawal-notice-to-ai-co-author>
- 18 Miriam Buiten, Alexandre de Streel, Martin Peitz, "The Law and Economics of AI Liability," *Computer Law & Security Review* 48, no. 105794 (2023), <https://www.sciencedirect.com/science/article/pii/S0267364923000055>
- 19 Dane Bottomley and Donrich Thaladar, "Liability for Harm Caused by AI in Healthcare: An Overview of the Core Legal Concepts," *Frontier in Pharmacology* 14, (2023), <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10755877/>
- 20 Mireille Hildebrandt, *Law for Computer Scientists and Other Folk* (Oxford University Press, 2020), Chapter 9, <https://lawforcomputerscientists.pubpub.org/pub/4swyxhx5/release/5>
- 21 Joy Buolamwini, "Unmasking the Bias in Facial Recognition Algorithms," *MIT Sloan*, December 13, 2023, <https://mitsloan.mit.edu/ideas-made-to-matter/unmasking-bias-facial-recognition-algorithms>
- 22 Carmen Niethammer, "AI Bias Could Put Women's Lives At Risk - A Challenge For Regulators," *Forbes*, May 02, 2020, <https://www.forbes.com/sites/carmenniethammer/2020/03/02/ai-bias-could-put-womens-lives-at-risk-a-challenge-for-regulators/?sh=61df083e534f>
- 23 Michael G. Jacobides, Stefano Brusoni, and Francois Candelon, "The Evolutionary Dynamic of the Artificial Intelligence Ecosystem," *Strategy Science* 6, no. 4 (2021): 412-435, <https://doi.org/10.1287/stsc.2021.0148>.
- 24 Luis Luna-Reyes et al., "Exploring the Relationships between Dynamic capabilities and IT governance: Implications for local governments," *Transforming Government: People, Process and Policy* 14, No. 2 (2020), <https://www.emerald.com/insight/content/doi/10.1108/TG-09-2019-0092/full/html>
- 25 Charles Jones and Christopher Tonetti, "Non-Rivalry and Economics of Data," *American Economic Review* 110, no. 5 (2020): 2819-58, <https://www.aeaweb.org/articles?id=10.1257/aer.20191330>.
- 26 Saffron Huang and Divya Siddharth, "Generative AI and the Digital Commons," *The Collective Intelligence Project*, Working Paper, 2024, <https://cip.org/research/generative-ai-digital-commons>

- 27 Jacobides, Brusoni and Candelon, “The Evolutionary Dynamic of the Artificial Intelligence Ecosystem”
- 28 Alan Chan, Herbie Bradley, Nitarshan Rajkumar, “Reclaiming the Digital Commons: A Public Data Trust for Training Data” (paper presented at 2023 AAAI/ACM Conference on AI, Ethics, and Society, New York, USA, 2023).
- 29 Jacobides, Brusoni, and Candelon, “The Evolutionary Dynamic of the Artificial Intelligence Ecosystem”
- 30 Dario Maisto, “Data Sovereignty Battles Continue To Dominate The European Public Cloud Market,” *Forrester*, November 18, 2022, <https://www.forrester.com/blogs/data-sovereignty-battles-continue-to-dominate-the-european-public-cloud-market/>; Anirban Ghoshal, “Microsoft in Talks over Cloud Licensing Complaint in the EU,” *CIO*, February 08, 2024, <https://www.cio.com/article/1306613/microsoft-in-talks-over-cloud-licensing-complaint-in-the-eu.html>
- 31 Ayanna Howard, “Real Talk: Intersectionality and AI”
- 32 Mary Clare Jalonick and Matt O’Brien, “Tech Industry Leaders Endorse Regulating Artificial Intelligence at Rare Summit in Washington,” *The Associated Press*, September 14, 2023, <https://apnews.com/article/schumer-artificial-intelligence-elon-musk-senate-cfcfb1067d68ad2f595db7e92167943c>
- 33 Samir Saran, Flavia Alves, Vera Songwe, “Technology: Taming – and Unleashing – Technology Together,” *Observer Research Foundation*, January 16, 2024, <https://www.orfonline.org/research/technology-taming-and-unleashing-technology-together>
- 34 Mohamed Elbashir, “EU AI Act sets the stage for global AI governance: Implications for US companies and policymakers,” *Atlantic Council*, April 22, 2024, <https://www.atlanticcouncil.org/blogs/geotech-cues/eu-ai-act-sets-the-stage-for-global-ai-governance-implications-for-us-companies-and-policymakers/>
- 35 “In an AI-Driven Digital Economy, How can Developing Countries Keep Up?” *UNCTAD*, December 08, 2023, <https://unctad.org/news/ai-driven-digital-economy-how-can-developing-countries-keep>
- 36 Jesse Klemptner, Christian Rodriguez, and Dale Swartz, *A Rising Wave of Tech Disruptors: The Future of Defense Innovation?*, McKinsey, 2024, <https://www.mckinsey.com/industries/aerospace-and-defense/our-insights/a-rising-wave-of-tech-disruptors-the-future-of-defense-innovation#/>; Shweta Surender, “Defense Industry Outlook: Emerging Defense Opportunities in 2024,” *Markets and markets*, February 08, 2024, <https://www.marketsandmarkets.com/blog/AD/Defense-Industry-Outlook>
- 37 “A Rising Wave of Tech Disruptors: The Future of Defense Innovation?”
- 38 Sarwant Singh, “Why The Defense Industry Outlook Is So Strong,” *Forbes*, March 11, 2024, <https://www.forbes.com/sites/sarwantsingh/2024/03/11/why-the-defense-industry-outlook-is-so-strong/?sh=60a3bc87a7a1>
- 39 Singh, “Why The Defense Industry Outlook Is So Strong”
- 40 Singh, “Why The Defense Industry Outlook Is So Strong”
- 41 Singh, “Why The Defense Industry Outlook Is So Strong”
- 42 Sigal Samuel, “Alexa, are you making me sexist?,” *Vox*, June 12, 2019, <https://www.vox.com/future-perfect/2019/6/12/18660353/siri-alexa-sexism-voice-assistants-un-study>
- 43 Samuel, “Alexa, are you making me sexist?”
- 44 Bedbible Research Centre, “Sex Robot Industry [New 2024 Data],” *Bedbible.com*, May 01, 2024, <https://bedbible.com/sex-robot-industry-market-size-technology-ai-user-sentiment-statistics/>



- 45 Chantal Cox-George and Susan Bewley, “I, Sex Robot: the health implications of the sex robot industry,” *BMJ Sexual & Reproductive Health* 44, no. 3 (2018)
- 46 Pallab Ghosh, “Sex robots may cause psychological damage,” *BBC*, February 15, 2020, <https://www.bbc.com/news/science-environment-51330261>
- 47 Jacobides, Brusoni, and Candelon, “The Evolutionary Dynamic of the Artificial Intelligence Ecosystem”
- 48 Barbara van Schewick, *Internet Architecture and Innovation* (MIT Press, 2010).
- 49 Jacobides, Brusoni, and Candelon, “The Evolutionary Dynamic of the Artificial Intelligence Ecosystem”
- 50 Jeremy Kahn, “AI Will Change the World. But that Doesn’t Mean Investors Will Get Rich in the Process,” *Fortune*, April 23, 2024, <https://fortune.com/2024/04/23/ai-foundation-models-llms-money-loser-for-investors-airlines-air-street/>
- 51 Jacobides, Brusoni, and Candelon, “The Evolutionary Dynamic of the Artificial Intelligence Ecosystem”
- 52 K V Kurmanath, “Google Will not Claim Ownership of AI-Generated Content,” *The Hindu BusinessLine*, April 22, 2024, <https://www.thehindubusinessline.com/info-tech/google-will-not-claim-ownership-of-ai-generated-content/article68091327.ece>
- 53 Anais Rességuier and Rowena Rodrigues, “AI Ethics Should not Remain Toothless! A Call to Bring Back the Teeth of Ethics,” *Big Data & Society* 7, no. 2 (2020); Karen Hao, “In 2020, Let’s Stop AI Ethics-Washing and Actually do Something,” *MIT Technology Review*, December 27, 2019, <https://www.technologyreview.com/2019/12/27/57/ai-ethics-washing-time-to-act/>; Brent Mittelstadt, “Principles Alone Cannot Guarantee Ethical AI,” *Natural Machine Learning* 1 (2019).
- 54 Anulekha Nandi, “The first international AI treaty: Progress with caveats,” Observer Research Foundation, May 22, 2024, <https://www.orfonline.org/expert-speak/the-first-international-ai-treaty-progress-with-caveats>
- 55 Nandi, “The first international AI treaty: Progress with caveats”
- 56 United Nations, *General Assembly Adopts Landmark Resolution on Artificial Intelligence*, United Nations News, 2024, <https://news.un.org/en/story/2024/03/1147831>
- 57 Jessica Carlo, Kalle Lyytinen, and Richard Boland, “Systemic Risk, IT Artifacts, and High Reliability Organizations: A Case of Constructing a Radical Architecture,” *Sprouts Working Papers on Information Systems* 4, no. 4 (2004).
- 58 Kalle Lyytinen, Lars Mathiassen, and Janne Ropponen, “Attention Shaping and Software Risk: A Categorical Analysis of Four Classical Risk Management Approach,” *Information Systems Research* 9, no. 3 (1998); Key Issues: Risk Based Approach,” EU AI Act, 2024, <https://www.euaiact.com/key-issue/3>
- 59 Julia Black and Robert Baldwin, “Really Responsive Risk-Based Regulation,” *Law & Policy* 32, no. 2 (2010), <https://onlinelibrary.wiley.com/doi/10.1111/j.1467-9930.2010.00318.x>
- 60 Anne Rouse, “The Governance Implications When it is Outsourced,” in *Information Technology Governance and Service Management: Frameworks and Adaptations*, ed. Aileen Cater-Steel (New York: Information Science Reference, 2009), 285-296.
- 61 OECD, “Technology governance and the innovation process,” in *OECD Science, Technology and Innovation Outlook 2018: Adapting to Technological and Societal Disruption* (Paris: OECD Publishing, 2018), [https://doi.org/10.1787/sti\\_in\\_outlook-2018-15-en](https://doi.org/10.1787/sti_in_outlook-2018-15-en).
- 62 John Braithwaite, “The Essence of Responsive Regulation” (Fasken Lecture, University of British Columbia, September 21, 2010).

- 63 Dynamic regulatory capabilities is drawn from the notion of dynamic capabilities by David Teece, “Technological Innovation and the Theory of the Firm: The Role of Enterprise-Level Knowledge, Complementarities, and (Dynamic) Capabilities,” in *Handbook of Economics of Innovation Vol. 1*, ed. Bronwyn Hall and Nathan Rosenberg (North Holland, 2010).
- 64 Philip Selznick, *The Moral Commonwealth: Social Theory and the Promise of Community* (Berkeley, CA: University of California Press, 1992).
- 65 John Braithwaite, “Responsive Regulation and Developing Economies,” *World Development* 34, no. 5 (2006): 884-896.
- 66 “Forty-two countries adopt new OECD Principles on Artificial Intelligence”, *OECD*, May 22, 2019, <https://www.oecd.org/science/forty-two-countries-adopt-new-oecd-principles-on-artificial-intelligence.htm>
- 67 NITI Aayog, *National Strategy for Artificial Intelligence*, 2018, <https://www.niti.gov.in/sites/default/files/2023-03/National-Strategy-for-Artificial-Intelligence.pdf>; NITI Aayog, *Responsible AI: Approach document for India*, 2021, <https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf>; NITI Aayog, *Responsible AI: Adopting the Framework – A Use Case Approach on Facial Recognition Technology*, 2022, [https://www.niti.gov.in/sites/default/files/2022-11/Ai\\_for\\_All\\_2022\\_02112022\\_0.pdf](https://www.niti.gov.in/sites/default/files/2022-11/Ai_for_All_2022_02112022_0.pdf)
- 68 Samir Saran, Flavia Alves, Vera Songwe, “Technology: Taming – and Unleashing – Technology Together”

## **About the Authors**

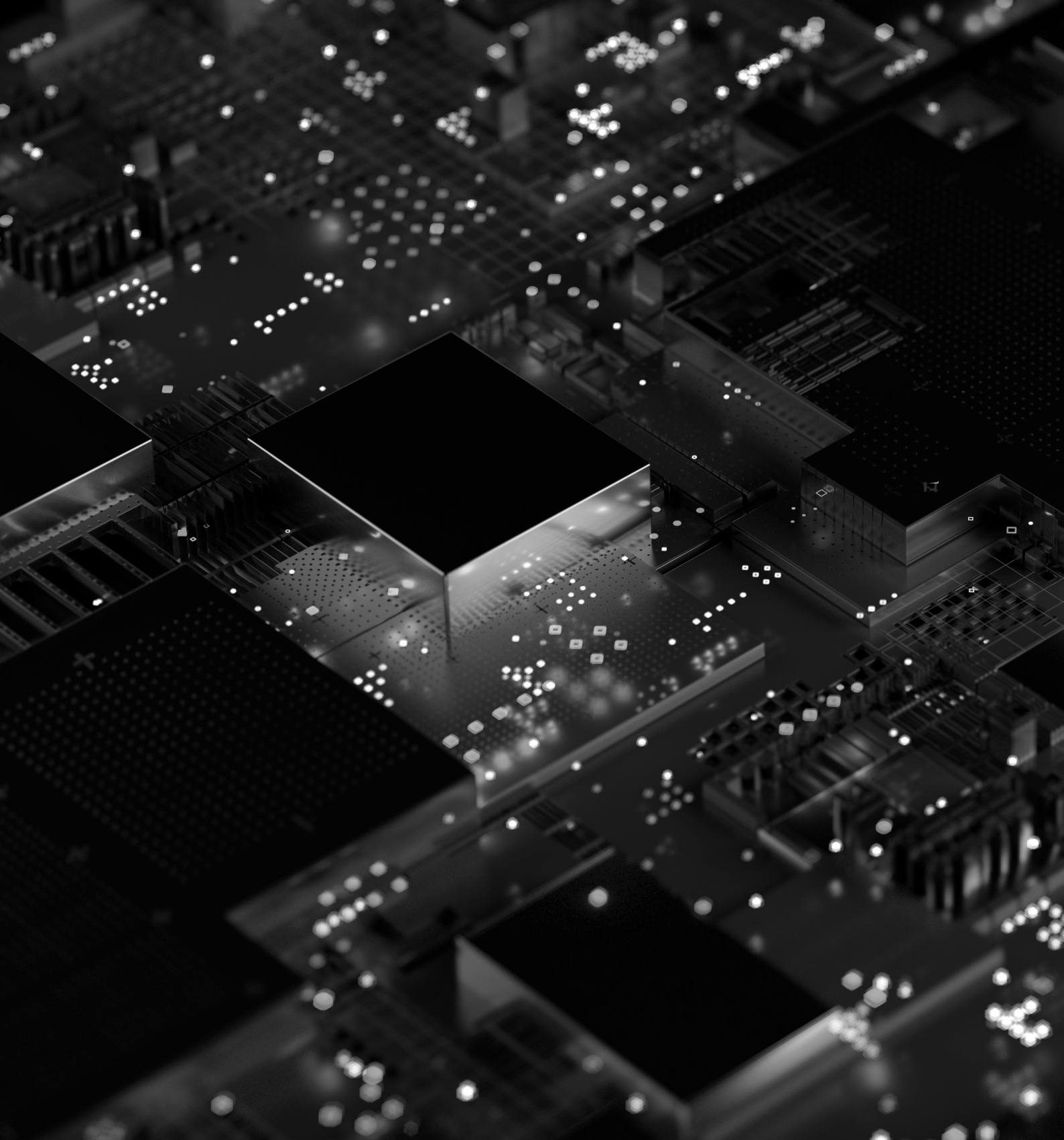
**Samir Saran** is President, ORF.

**Anulekha Nandi** is Fellow, Technology, Economy, and Society, ORF.

**Sameer Patil** is Senior Fellow, Centre for Security, Strategy, and Technology and Deputy Director, ORF Mumbai.

*Cover photo: Getty Images/Ludovic Marin*

*Back cover image: Getty Images/Fabrice Coffrini*



**Ideas . Forums . Leadership . Impact**

**20, Rouse Avenue Institutional Area,  
New Delhi - 110 002, INDIA**  
**Ph. : +91-11-35332000. Fax : +91-11-35332005**  
**E-mail: [contactus@orfonline.org](mailto:contactus@orfonline.org)**  
**Website: [www.orfonline.org](http://www.orfonline.org)**