Centre for International
Governance Innovation

CIGI Papers No. 302 — September 2024

Philanthropy's Urgent Opportunity to Create the Interim International AI Institution

David Evan Harris and Anamitra Deb

Centre for International
Governance Innovation

CIGI Papers No. 302 — September 2024

Philanthropy's Urgent Opportunity to Create the Interim International AI Institution

David Evan Harris and Anamitra Deb

About CIGI

The Centre for International Governance Innovation (CIGI) is an independent, non-partisan think tank whose peer-reviewed research and trusted analysis influence policy makers to innovate. Our global network of multidisciplinary researchers and strategic partnerships provide policy solutions for the digital era with one goal: to improve people's lives everywhere. Headquartered in Waterloo, Canada, CIGI has received support from the Government of Canada, the Government of Ontario and founder Jim Balsillie.

À propos du CIGI

Le Centre pour l'innovation dans la gouvernance internationale (CIGI) est un groupe de réflexion indépendant et non partisan dont les recherches évaluées par des pairs et les analyses fiables incitent les décideurs à innover. Grâce à son réseau mondial de chercheurs pluridisciplinaires et de partenariats stratégiques, le CIGI offre des solutions politiques adaptées à l'ère numérique dans le seul but d'améliorer la vie des gens du monde entier. Le CIGI, dont le siège se trouve à Waterloo, au Canada, bénéficie du soutien du gouvernement du Canada, du gouvernement de l'Ontario et de son fondateur, Jim Balsillie.

Credits

Research Director, Transformative Technologies Tracey Forrest
Director, Program Management Dianna English
Program Manager Erin Chreptyk
Publications Editor Lynn Schellenberg
Publications Editor Christine Robertson
Graphic Designer Sami Chouhdary

67 Erb Street West
Waterloo, ON, Canada N2L 6C2
www.cigionline.org

Table of Contents

About the Authors

David Evan Harris is a CIGI senior fellow, a Chancellor's Public Scholar at the University of California, Berkeley, and a faculty member at the Haas School of Business, where he teaches courses on artificial intelligence (AI) ethics, social movements and social media, civic technology, futures thinking and scenario planning. David is also a senior research fellow at the International Computer Science Institute, a senior policy adviser at the California Initiative for Technology and Democracy, and a senior advisor for AI and Elections to the Brennan Center for Justice at NYU. He was named to Business Insider's AI 100 list in 2023.

David previously worked as a research manager at Meta (formerly Facebook) on the responsible AI and civic integrity teams. His writings and commentary have been featured by *The Wall Street Journal*, *The Washington Post*, CNN, *The Guardian*, BBC, Associated Press, Bloomberg, *IEEE Spectrum*, *The Atlantic*, Tech Policy Press and Adbusters. David has advised the White House, the US Congress, the European Union, the United Nations, the North Atlantic Treaty Organization and the California State Legislature about technology policy. He has conducted research, studied or given talks in person in 37 countries and speaks fluent English, Portuguese and Spanish, as well as intermediate French.

Anamitra Deb currently serves as managing director of responsible tech at Omidyar Network, a philanthropic organization whose mission is to bend the arc of the digital revolution toward shared power, prosperity and possibility. He oversees teams working on technology platform accountability and regulation; privacy, data governance and online trust and safety; better tech cultures and products; responsible and open-source tech; and guardrails for emergent technologies such as generative AI.

Anamitra has authored several reports in responsible tech, market-based solutions, impact investing and strategic philanthropy, and lectured on these topics at the University of California, Berkeley, and Georgetown University, among others. His writings and perspectives have been featured in major publications, including *The Economist*, *The Washington Post*, *Wired*, *Barron's*, Bloomberg, *The Wall Street Journal*, *Financial Times* and *The Times of India*.

Anamitra is a senior fellow at the Munk School of Global Affairs and Public Policy at the University of Toronto. He sits on several advisory boards including Consumer Reports' Innovation Council and the Center for Media Engagement at the University of Texas at Austin.

Acronyms and Abbreviations

AI	artificial intelligence
AIG-CSAM	AI-generated child sexual abuse material
CEN	European Committee for Standardization
CENELEC	European Electrotechnical Committee for Standardization
CNN	Cable News Network
CSAM	child sexual abuse material
GPAI	general-purpose AI
IIAII	Interim International AI Institution
NAIRR	National Artificial Intelligence Research Resource
NCII	non-consensual intimate imagery
OECD	Organisation for Economic Co-operation and Development
SDGs	Sustainable Development Goals
UNF	United Nations Foundation
UNFIP	United Nations Fund for International Partnerships
WHO	World Health Organization

Executive Summary

The rapid development of new AI technologies has outpaced the ability of regulators in most parts of the world to put rules in place that govern their use. Without regulation of AI, its benefits are likely to flow to the few, while the many risks it poses and the harms it has already wrought will be borne by society, and disproportionately so by already vulnerable communities. Many of AI's problems are inherently global, which means that if regulation takes place in a loose patchwork, solutions will be evasive: certain AI products, services, practices or tools outlawed in one part of the world may still be available through companies located in other jurisdictions. The authors argue that in the context of the urgent need for truly meaningful regulation of AI, philanthropy has an opportunity to quickly leverage its stores of uniquely public-interest-bound "risk capital" to create the Interim International AI Institution. While numerous efforts are under way to start conversations and study what a possible international or intergovernmental AI governance body might look like, this proposal uniquely suggests simply putting forth the funding and prototyping the organization by starting on the work today. The critically important work of this institution would include coordinating conversations among governments around the world that are now developing potentially incompatible AI governance regulations in parallel; establishing best practices and norms for AI governance; bringing together a critical mass of technical, legal, policy and social science expertise; and transparently sharing the fruits of its rapid and iterative AI governance prototyping efforts.

Introduction: The Pacing Problem and the Philanthropic Solution

Background: AI on the Horizon

The pace of AI[1] innovation is dizzying. Even industry insiders struggle to keep up with the daily announcements of new technology developments and the breakthroughs regularly on the horizon. Around the world, observers are expressing a mix of, on the one hand, ebullience about the societal value of AI, and on the other, widespread concern about not only the current harms stemming from AI, but also those that might arise in the future (De Tena et al. 2023; Orth 2023).

The hoped-for benefits of AI remain a driving force for the industry's continued development. Proponents of AI development cite many potential benefits, ranging from product innovations such as digital assistants and self-driving cars to monumental changes in society, including solutions to climate change and a cure for cancer. That being said, AI experts broadly agree that with AI's benefits come risks (Grace et al. 2024, 12). These include the use of AI in surveillance, manipulation, propaganda, creation of non-consensual intimate imagery (NCII), violation of privacy and development of dangerous weapons that could pose catastrophic risks to society (Hendrycks, Mazeika and Woodside 2023; Lakatos 2023). Others point out that the harms of AI are already here, citing AI's role in discrimination in criminal sentencing (Larson et al. 2016), as well as in access to health care, housing, lending and jobs (Backman 2023; McIlwain 2020; Akselrod and Venzke 2023).

An unexpected twist that distinguishes the business of AI from many other industries is that the CEOs of nearly all the major companies (Zakrzewski, Lima-Strong and DiMolfetta 2023) involved in

1 In this paper, we refer to AI broadly, including the ranking and recommendation systems that power social media platforms as well as generative AI technologies. Specifically, we use the updated definition of AI from the Organisation for Economic Co-operation and Development (OECD) in 2023: "An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment" (Russell, Perset and Grobelnik 2023).

AI have themselves called for regulation of the technology — as these companies face market pressures to release new products more quickly than their competitors, they are simultaneously publicly worrying that the frenetic pace of these releases could pose risks to society if there are not rigorous AI safeguards and regulation in place (McCracken 2023). In a May 2023 congressional hearing, OpenAI CEO Sam Altman stated, "I think if this technology goes wrong, it can go quite wrong," and said that OpenAI would "want to work with the government to prevent that from happening" (Kang 2023). In a September 2023 congressional committee hearing, other major technology business leaders, including Elon Musk and Mark Zuckerberg, also agreed that the government needs to play some role in regulating AI (Wong et al. 2023).

Governments Fall Short

In the face of the aforementioned calls for regulation, elected officials from around the world are working hard to establish regulations for AI. Unfortunately, their efforts have been stymied by two major challenges. The first is the geographic flexibility of AI companies who can simply relocate their way out of regulation by ceasing to do business in jurisdictions that choose to impose regulations, or at a minimum, threaten to do so. Sam Altman demonstrated this corporate capability when he threatened to remove OpenAI's products from Europe if the EU AI Act required AI companies to disclose the use of copyrighted materials in their training data. The European Union's Commissioner for the Internal Market, Thierry Breton, called Altman's bluff, and later Altman recanted (Field 2023; Chee 2023). Nevertheless, this dynamic puts significant pressure on regulators who do not want their work to be seen as an "innovation killer" (O'Reilly 2023), and the only real solution is for governments to band together in regulating AI. The second major challenge is the "pacing problem," described in greater detail below, which is a known issue wherein governments struggle to make regulations at the same pace as technological development.

The United States Congress — the most powerful democratically elected body with the potential to meaningfully regulate the AI industry — remains an institution designed to move slowly and deliberately (S. Turner 2016). While draft bills on AI have been circulated, it remains uncertain, at best, whether any of the comprehensive approaches will become law any time soon (Covington & Burling LLP 2023). Without federal legislative action on AI, any other approach will be limited in its ability to stop the current "race to the bottom" on AI ethics and safety (Harris 2023a). State- and municipal-level legislation can be at least partially circumvented through corporate relocation (as discussed above), and the force of public opinion has only so much power. Left to the economic pressures that incentivize the current rapid development of AI technologies, it appears likely that AI companies in the United States will continue to minimize expenditures on ethics and safety while they maximize the development of new system capabilities.

In October 2023, the White House, seeking to make progress on AI regulation without needing congressional approval, issued an executive order to signal the administration's priorities and to urge several agencies and departments to take meaningful action on AI at the federal level (Harris 2023b). Despite striking a balance between the interests of industry and the public interest, this approach has inherent limitations: executive orders can be easily reversed by subsequent administrations and lack the congressional power to allocate funds (Thrower 2021).

For those interested in ensuring that AI brings the world maximum benefit and minimum harm, the frustrating lack of durable and meaningful progress in Washington leaves us wondering where to look. Around the world, cities, states, countries and regional organizations are beginning to create their own AI legislation,[2] creating a weak patchwork of laws that make up the emerging regulatory environment for AI.

The initial elements of this patchwork came from China, as the first nation to implement firm AI regulations, starting in 2017. Unfortunately, given the country's long history of using technology for surveillance, censorship and control of its population, while some of its laws prohibiting deepfakes and curbing the power of AI developers seem reasonable, the overall approach does not provide an imitable model for democratic nations (Heath 2023). There are key instances of overlap between Chinese and democratic interests, such as China's novel provisions on "Deep Synthesis Technologies," which aim to regulate each step of

2 See Johnson (2023) and https://iapp.org/resources/article/global-ai-legislation-tracker/.

the deepfake process, from deepfake generation to sharing. While there may be disagreements over certain AI policies, it is clear that some of the most significant threats posed by AI, such as the use of deepfakes, serve as a common ground on which democratic and non-democratic countries might collectively build regulation (Sheehan 2023).

The European Union, with its 2024 AI Act,[3] followed China and has now emerged as the most significant player among democratically governed nations on AI regulation (Chan 2023). The AI Act places bans on numerous uses of AI, including "cognitive behavioral manipulation" — a broad term for uses that interpret behaviours and preferences with the intent to influence our decisions. The bans also include the "untargeted scraping of facial images from the internet or CCTV footage," a practice already used by some companies that sell databases used for surveillance; "emotion recognition in the workplace and educational institutions," which could be used by companies to discipline, rank or micromanage employees; "social scoring," a surveillance tool used in China to rate individuals on everyday activities and award (or deny) them social credit;[4] "biometric categorization," a practice that uses characteristics such as skin tone or facial structure to infer gender, sexual orientation or even the likelihood of committing a crime; and "some cases of predictive policing for individuals," which have already been shown to have racially discriminatory impacts (Harris 2023d).

The EU AI Act also regulates "General-Purpose AI systems" (GPAI), which have great potential to be used for both good and harm. Although the law only applies directly to AI used in Europe, law has significant extraterritorial implications, in that AI developers doing business in Europe, even if not based there, will have to comply with aspects of the law in ways that are likely to shape their business practices and products around the world. These include provisions that require AI developers to produce risk assessments of their AI systems, to take precautions to mitigate those risks and to share information about their energy consumption.

However, even strong proposals from the world's leading countries may not be enough to guarantee the safe development of AI technology globally. Relying on the "Brussels effect," a term for the phenomenon whereby regulations adopted in Europe naturally become adopted as de facto global standards, may not be enough to ensure the safe proliferation of AI (Bradford 2020). Given the scale of the global business, political and individual self-interests at play, the example set by the EU AI Act may not have the widespread impact that the Brussels effect implies. One category of harm in particular that cannot be stopped by a weak patchwork of laws is the misuse of powerful AI systems by "bad actors," who may range from vengeful individuals creating devastating NCII, to scammers cloning the voices of victims' relatives, to intelligence agencies and digital mercenaries using AI for coordinated manipulation and misinformation efforts, to interference in elections, to developing lethal autonomous weapons, to the deployment of powerful cyber weapons.

Absent any meaningful regulation, companies including Meta, Stability AI, Hugging Face, Mistral AI, EleutherAI and the Technology Innovation Institute have chosen to distribute their AI systems in ways that can be easily misused by bad actors. These companies are effectively racing against each other to release ever more powerful "open" and "open-source" AI systems, referred to herein as "unsecured" AI systems, to signify that their model weights have been publicly released in ways that facilitate repurposing — which can in some cases be good, but in others can impose significant risks on society (Harris 2023c). While these unsecured models aren't the only models susceptible to misuse, online interfaces to secured AI (systems offered through hosted web or application programming interfaces) offer opportunities to stop bad actors from accessing and abusing high-risk AI on a large scale. While hackers have found ways to circumvent safety features of secured AI systems, the developers of these systems are able to patch vulnerabilities once they have been discovered and also to limit the rates of usage of their systems. These types of security interventions are not possible for unsecured AI systems, which can be downloaded, fine-tuned to facilitate abuse and run in secret. No security measure is ever perfect, as hackers and spies could potentially

3 EU, *Regulation 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)*, [2024] OJ, L 12.7.

4 A concerning example of this includes Chinese journalist Liu Hu, who was blocked from buying plane tickets, property or taking loans. There was no formal government notice of the restrictions or an appeal process, with the restrictions being thought to be the result of Hu's anti-corruption reporting and tweets against the government (Zhao 2018).

steal model weights of secured systems, but this should not be taken to mean that security measures ought to be abandoned completely. Regulation will only become more crucial as time passes, with higher capability models from both open and closed sourcing becoming even more risky if used by malign actors (Seger et al. 2023, 12).

One telling illustration of these vulnerabilities can be seen in an announcement from Microsoft and OpenAI that hackers from China, Iran, North Korea and Russia were caught using their AI systems to improve their cyberattack techniques. Once the hackers have been caught, the companies can block them and develop more sophisticated ways of detecting abuses based on their usage patterns. The same threat actors, however, could very easily migrate their efforts to the unsecured systems provided for download by the companies listed above, and likely never be detected or stopped from abusing them (Satter 2024).

Today, this particular set of companies releasing unsecured AI systems is based in Europe, the United Arab Emirates, the United Kingdom and the United States. Of these companies, the EU AI Act applies most directly to Mistral AI, headquartered in France. Even though the act has yet to come into force, it appears to have had an impact on the company's release strategy. Co-founded by France's former digital minister Cédric O, Mistral AI fiercely resisted the EU AI Act's regulations on GPAI and sought to add specific exemptions for "open-source" AI to the act (Wanat 2023; Chatterjee and Volpicelli 2023). A few months after the EU AI Act was confirmed to apply to GPAI, with only limited exemptions for less powerful "open-source" models, the company announced a partnership to distribute a new, secured model, "Mistral Large," in partnership with Microsoft (Leprince-Ringuet 2024). However, with the weak patchwork of regulations that we have today, a company in this position could simply relocate to a more permissive jurisdiction and continue to develop unsecured AI systems, albeit without commercial access to the European Union market.

What we see here is a perfect illustration of the "pacing problem," where the pace of technology development outstrips the pace of governmental regulation (L. Downes 2009). This problem is evident in a wide variety of technical fields, including stem cell research, genetic testing, synthetic biology, nanotechnology, neuroscience, driverless cars, human cloning and geoengineering (Kuokkanen and Yamineva 2013; Marchant 2011), but the pace of AI development today seems even faster than the pace of development in these other areas.

Enter Philanthropy

Faced with such an impasse, international networks of academics have begun working with civil society organizations such as Access Now, the Algorithmic Justice League, Amnesty International, the Center for AI and Digital Policy, the Center for the Advancement of Trustworthy AI, the Centre for International Governance Innovation (CIGI), the Future of Life Institute and many others, in an effort to advance AI governance efforts around the world.[5]

Organized philanthropy, used here to refer to charitable foundations and their grant-making initiatives, has begun to commit significant financial resources to these efforts. Based on our review of publicly available information from the Foundation Directory,[6] we estimate that more than $300 million have been granted by US private foundations to AI programs between 2018 and 2023, with roughly one-third of that total going to AI governance and policy efforts.[7]

Strategic philanthropists today have an outsized opportunity to bridge the gap between the runaway speed of AI technological development and the multi-layered deliberative processes that could take years — or even decades — before a stable intergovernmental AI regulatory body is established.

The authors of this paper argue herein for large-scale philanthropic investment to create a new Interim International AI Institution (IIAII, pronounced "aye-aye") that could act as a stopgap measure to facilitate international collaboration on AI governance. With organized philanthropy's support, funds could be allocated to establish this prototype AI governance body

5 See www.aiethicist.org/ai-organizations; Belfield (2020, 16).

6 See https://fconline.foundationcenter.org/.

7 Other categories that we assigned to grants included AI and medicine/science, AI ethics, AI and climate change, AI education, AI fairness and inclusion, and AI safety/existential risk. It is not appropriate to consider this an exhaustive review or anything more than a rough approximation of a floor on funding, due to potential significant gaps in the Foundation Directory database, delays in reporting, and the difficulty of assigning categories to grants that often blurred the lines between categories or fell into multiple categories. Corporate giving coming directly from a company was also not included in these totals.

immediately, without the lengthy process of securing commitments from governments around the world. An institution like this one, if thoughtfully constructed, could help build the capacity for international collaboration on AI governance and support policy makers around the world already looking to collaborate with one another — in absence of such an institution. This institutional prototype would bring together policy experts, social scientists and AI researchers with diverse disciplinary backgrounds, facilitating the attainment of technical and policy consensus where possible, and clearly identifying areas where compromise will be necessary.

Fortunately, there are historical models of international agreements and related bodies from other industries that have been studied as models for AI governance, including the International Civil Aviation Organization (Trager et al. 2023, 19), the International Atomic Energy Agency (Nichols 2023), the Intergovernmental Panel on Climate Change (Suleyman and Schmidt 2023), the Financial Action Task Force and the International Telecommunication Union (Ho et al. 2023, 9).

Some scholars describe philanthropic resources as society's "risk capital" — the funds held by wealthy individuals or their foundations that can often be deployed quickly and with greater risk tolerance than government investments (Bosire 2020; Buck 2020). Philanthropic wealth is never apolitical, and it will be critical to put checks on the power of any funders contributing to the IIAII. Global civil society — of which private philanthropy is a part (McGuigan and Bass 2022) — has an important role to play in shaping the global governance of AI and should be called on to guide the path forward by advising philanthropists on the investments recommended here.

Historical Antecedent: Ted Turner's Billion-Dollar UN Gift

The idea of strategic philanthropic engagement in support of international governance institutions is not new. The most salient and perhaps inspiring case study is Ted Turner's decision to donate $1 billion to the United Nations in 1997. Although this donation was mostly known for its historic size, it also had important structural implications for UN funding more broadly (United Nations 2006). In 1997, the United Nations faced criticism of its administrative spending, which ultimately led it to reform its operations the following year (*The New York Times* 1997; US Department of State 1997). Leading the charge for these critiques, the US government began to withhold more than $1 billion in membership dues in protest of the United Nations' inefficiencies (Crossette 1997).

In the face of these criticisms, Ted Turner announced his $1 billion donation, originally aimed to "erase the debt" the US government owed to the United Nations. Turner intended the donation to allow the United Nations to continue operating specific programs that had been frozen due to lack of funding (C. Turner 1997). By stepping in when government funding was falling short, Turner demonstrated that intergovernmental initiatives, and the United Nations in particular, could be funded by sources other than government bodies in a way that had positive geopolitical ramifications. Moreover, the realized effects of Turner's donation also helped to further facilitate civil society and private sector engagement with the United Nations in the years that followed.

With his donation, Turner created the United Nations Foundation (UNF), a separate organization founded by Turner with an initial objective of funnelling the large donation to "UN causes" (Cable News Network [CNN] 1997). Kofi Annan, former UN Secretary-General, credited the founding of the UNF as the moment that the United Nations became a "partnership organization," an important transition that greatly increased its capabilities (United Nations 2006). By acting as a funding body for the United Nations, the UNF created a mechanism for companies and individuals to donate money to UN causes, which has since

become an important part of the United Nations' sustainable development efforts.[8] The UNF has also inspired and directly channelled hundreds of millions of dollars in additional donations from individuals, corporations, governments and non-governmental organizations to UN agencies (ibid.). In this way, the UNF serves as both a legal and a financial mechanism and could be seen as a historical antecedent for the proposed IIAII, demonstrating the potential for philanthropic contributions to meaningfully change the course of international governance institutions.

Of course, there are also reasons to be skeptical of such mechanisms — ideally, the United Nations would never have found itself in the position it did in 1997, when the United States defaulted on its debt. But at that historical moment, Turner's intervention played a positive role for the organization and leveraged additional resources. The United Nations took steps, however, to make sure that the application of UNF contributions would be carefully governed by UN officials. This governance takes place through the United Nations Fund for International Partnerships (UNFIP), which today sits within the UN Office for Partnerships, as well as through the UN-UNF Joint Coordination Committee. UNFIP's advisory board is chaired by the Deputy Secretary-General.[9] The careful structuring of this relationship could be studied in support of the design of a similar mechanism that might tie the IIAII to existing UN bodies or facilitate a planned transition to such a relationship in the future. Depending on the direction of UN efforts on AI governance, the IIAII as an institution could, pending approval of UN leadership, potentially even be donated in its entirety to the United Nations via the UNF and UNFIP.

The long-standing relationship between the Bill & Melinda Gates Foundation and the World Health Organization (WHO) is another example of a major philanthropic contributor dramatically increasing the capacity of an intergovernmental institution. According to Euronews, "In 2018–2019, the United States was the largest donor at $893 million, accounting for around 15 per cent of WHO's budget. The Gates Foundation came only second, with $531 million" (Carbonaro 2023). Based on information published on the Gates Foundation's own website, its contribution to the WHO from 1998 to 2023 appears to be well over $5 billion, not adjusted for inflation.[10]

When it comes to the governance of AI, the situation is different. There is currently no intergovernmental body that plays a role in AI governance. This presents a once-in-a-generation opportunity for one or more ambitious philanthropists to step up and use their financial resources to bridge the gap between the speed of AI governance and the speed of AI technology development.

How Philanthropy Is Positioned to Help

Philanthropy — along with civil society, policy makers and industry — has an urgent opportunity to push AI to fulfill its potential to deliver enormous benefits to society. It would be far from the first time that philanthropy has bridged the gap between public and private interests, having had core involvement in fields ranging from climate change (Betsill et al. 2022) to access to COVID-19 vaccines (Banco, Furlong and Pfahler 2022) to nuclear non-proliferation (Rubinson 2021). None of these efforts have transpired, however, without controversy (Morena 2023; Sklair and Gilbert 2022; Sparke and Levy 2022), and it is critical that philanthropists interested in boldly funding AI governance efforts learn from both the successes and the failures of these past efforts.

Notably, philanthropy even played a significant role in the birth of AI itself. In 1956, the Rockefeller Foundation gave a grant to support the Dartmouth Conference, a five-week-long gathering of researchers that has now become known as the birthplace of the modern notion of AI. In what might be one of the most historically significant grant proposals of all time, the proposal for the Dartmouth Conference contains the first documented use of the term *artificial intelligence* (Shubinski 2022). Today we have come full circle, to a moment when organized philanthropy must once again play a critical role in shaping the field of AI.

8 See https://unfoundation.org/what-we-do/initiatives/.

9 See https://unfoundation.org/who-we-are/our-financials/un-foundation-un-relationship/.

10 See www.gatesfoundation.org/about/committed-grants.

Reflecting on both current and historical examples, the authors propose a three-part framework for understanding how philanthropy can approach supporting the development of international AI governance in a manner that is inclusive and participatory; provides critical capital and support for innovation, speed and risk in public policy approaches; and supports the creation and sustenance of institutional infrastructure that can increase capacity and resilience in the digital technology ecosystem.

Ensure Inclusive Representation and Participation of Civil Society

One of philanthropy's best contributions has been to build the capacity of diverse, expert and timely coalitions in the face of major technological change, both to ensure that technology's benefits are distributed and democratic, and to mitigate its harms (Slaughter and Walker 2021). It is time to ask philanthropy to do this once more.

In today's technology and media industries, a small number of actors play an outsized role in decision making and value capture (Hutchinson 2022; Moore and Tambini 2022). This leads to a massive asymmetry in who benefits from technology and who bears the costs of both targeted harms (for example, scams and fraud, biased decisions) and diffuse harms (for example, disinformation, mental health, and so forth) (Malik et al. 2022; Robinson and Edwards 2024; Thakur and Hankerson 2021). The starting point for philanthropy should be that the power, value and decision making of technology cannot be concentrated in the hands of a few, be they corporations, governments or the wealthy.

One key philanthropic strategy in this domain is to ensure that individuals with diverse lived experiences and perspectives actively shape the design, deployment, monitoring and impact of AI. Philanthropy has played a critical role in ensuring that such individuals — and the organizations they lead or are affiliated with — are capitalized to make such contributions. In the United States, for instance, strategic philanthropists over the past several decades have tried to ensure that diverse champions, organizations and coalitions with expertise in emerging technology and their surrounding governance and accountability ecosystems are funded and supported. One example of this type of work is the Rockefeller Foundation's grant of $300,000 to Black in AI, a technology research organization, to enhance the representation of Black individuals in the field of AI. This initiative aims to cultivate a new network of Black scholars and engineers while combatting bias in AI (The Rockefeller Foundation 2022). Specifically, part of the funds were allocated to addressing issues of discrimination toward people of colour in AI facial recognition technology. Indeed, what might be broadly called the responsible technology (or public interest technology) movement has seen evolving battlegrounds, from net neutrality to open data and the right to information, to privacy and data governance, to competition and antitrust, and even to trust and safety (Omidyar Network 2022).

AI discourse today is influenced by the ongoing sparring between what is colloquially known as "AI ethics" and the "effective altruism" or "X-risk" movements; a set of (sometimes divergent) foundations and networks have undergirded each field (Agüera y Arcas 2023). We agree with calls from these divergent communities for recognition that their struggles are bound up with one another and that solving AI's present-day ethical harms is in many ways a precondition for addressing AI's longer-term risks (Kubzansky 2024).

Equally important is the role that philanthropy can play, and has played, in bringing together groups working on behalf of various interests — for example, labour, civil rights and justice, privacy and data governance, even climate — to be represented at the negotiating table on critical issues. Too often, the potential of rapidly advancing technologies (including but not limited to AI) to cause serious and widespread harm is recognized only after the fact. These harms are often disproportionately borne by some of the aforementioned interest groups who are not usually considered technology's stakeholders. By learning from past errors and supporting these key constituencies, philanthropy can play an important role in both staving off the downsides of technological advances and fostering broad-based coalitions that work together for the collective good.

Yet another way for philanthropy to play an important role in building inclusivity is to ensure that the responsible tech ecosystem has the capacity to speak out quickly as an early-warning system for unforeseen harms and unintended consequences. This capacity was visibly evident when ChatGPT broke consumer adoption records in 2022–2023 (Hu 2023). Since academic researchers and civil society organizations had been studying AI bias and fairness with philanthropic support for

years before the arrival of ChatGPT (Partnership on AI 2020, 9), they were prepared to make recommendations on what to measure, how to understand progress and warnings, and where to invest in disclosure and oversight — from audits, to red teaming,[11] to potential licensing and accountability metrics and authorities. Philanthropy has long supported experimentation and innovation in many of these areas, with the privilege of having both a longer-term time horizon than most other institutions and the ability, when needed, to take risks and move quickly.

In many ways, philanthropy fills in the gaps when trying to establish some parity between the big power players (corporations and governments) and those who are trying to raise the clarion call for better and improved stewardship (Ford Foundation 2023). Underlying all of these goals is a critical insight for the role of philanthropy: the entire project is, in a sense, about ensuring that the digital technology ecosystem itself centres humanity and societal well-being. This means embedding both long-term thinking and rapid-response capabilities that can be resilient and maneuverable to match the speed and evolution of any technology. Such an ecosystem would have characteristics that allow it to respond to both structural issues (for example, privacy and data governance, trust and safety, and so forth, which continue to be governed by foundational principles) and emergent issues (for example, confabulation and synthetic child sexual abuse material [CSAM], which may be unsavoury or illegal attributes of a particular technological advancement).

Providing Risk Capital for Public Policy Development

In philanthropic circles, the notion of supporting "experimental pilots" of solutions to social problems appears to be gaining popularity.[12] Philanthropists are also increasingly looking to the term *catalytic capital* as an expression of how they can take risks to fund projects that catalyze deeper societal change rather than address problems at a superficial level (Schwartz 2024). This has not always been the case. There is wide agreement (Knapp 2023; Law 2023) that the mixing of innovation and philanthropy was led by technology and media industry leaders such as Laurene Powell Jobs, Ted Turner, Melinda and Bill Gates, Pam and Pierre Omidyar and the X Prize Foundation.

In the field of AI, there is an opportunity to double down on these types of "high-risk, high-reward" projects (Buck 2020). A logical extension of these approaches is to test novel public policy approaches to rapidly advancing technologies such as generative AI.

Funding projects designed to build civil society "go-to" options for model evaluation and improvement tools will go a long way toward ensuring that we scale the best, most effective policy solutions possible. These initiatives have included testing AI's resistance to adversaries through red-teaming exercises, using algorithmic audits to evaluate bias and boost fair decision making, or integrating public rating systems for large language models. Data and Society's Algorithmic Impact Methods Lab and Dr. Rumman Chowdhury's work on "humane intelligence" are examples of these efforts.[13] Similarly, important work can be done to ensure that we support researchers seeking answers to the difficult questions about emergent areas, such as "How will we ensure that technology is used to augment human capabilities and not replace them?" and "How will our relationships evolve in the age of intelligent and communicative machines?"

Another role of philanthropy is to provide funding to leverage and publish lessons from both successes and failures to advance the public conversation about what works and why. For instance, in areas such as trust and safety, philanthropy has supported associations of fellows to ensure knowledge sharing and the development of best practices,[14] supported journals that supply innovation and pilot results,[15] and created fora for sharing knowledge about guideposts and diagnostic tools for better outcomes (Stanford Internet Observatory 2023).

Finally, philanthropy can help to accelerate consistent global standards and interparty negotiations on policies and regulations. Bringing harmonization and interoperability to the approach that governments around the world are taking to

11 See https://csrc.nist.gov/glossary/term/red_team.

12 See www.macfound.org/grantee/lever-for-change-10114942/ and www.bdmorganfdn.org/morgan-startup-grants.

13 See, respectively, https://datasociety.net/algorithmic-impact-methods-lab/ and www.humane-intelligence.org/.

14 See https://integrityinstitute.org/our-supporters.

15 For example, see https://tsjournal.org/index.php/jots/about.

AI is critical, and here the philanthropic sector can speed up ongoing diplomatic work. Consider, for example, organized philanthropy's funding of the UN Secretary-General's High-level Advisory Body on AI (Advisory Body on Artificial Intelligence 2023, 25). This body brings together a cross-disciplinary group of 38 global experts from all world regions to offer diverse perspectives and options on how to govern AI for humanity, including in support of the United Nations' Sustainable Development Goals (SDGs). Philanthropy's fast-turnaround support for efforts such as this allows action to be taken without the delays involved in awaiting member state contributions. This type of funding has the potential to assure that the public sector is not caught flat-footed on issues such as AI governance that require global coordination and rapid action.

Supporting Shared Infrastructure and New Institutions

Digital technology continues to advance faster than laws, regulations, policies, market incentives and societal norms can keep up (Marchant 2011). As a result, it has become an ongoing struggle to address both narrow and widespread harms, usually long after they have taken a heavy toll. As noted above, philanthropy can serve to bridge the lessons of the past with the emerging needs of today, as they have done to assist policy makers, regulators and business leaders for other complex and multipurpose technologies, including nuclear technology and biomedicine (Toma 2022).

To do this work, philanthropy must invest in the creation and sustenance of new governance infrastructure. Examples might include funding new capacities (for example, governmental ability to respond to threats through research, mobilization and advocacy) and human capital pathways (so that there is a threshold of dedicated expertise that can work in the field), as well as new fields of research and inquiry and new organizational homes. Philanthropy has done this before in areas as diverse as public interest technology and AI ethics (Ford Foundation 2023), impact investing (The Rockefeller Foundation 2021) and drug and vaccine delivery (NBC News 2005).

In AI, examples of such infrastructure already in development in civil society are the Center for the Advancement of Trustworthy AI, which focuses on providing governments with turnkey tools, training, consulting and best practices for AI regulation, and the Distributed AI Research Institute, which prioritizes independent, community-based research. In the public sector, philanthropy has supported efforts to consider the provision of AI technology in the public interest through proposals such as the National AI Research Resource, or NAIRR (which piloted in January 2024[16]), and CalCompute (envisioned as a publicly owned and operated cloud-computing cluster run by a consortium of universities and resesarch labs; see Sitaraman and Foster 2024).

The pace and the scale of AI's progress demand the creation of a new ecosystem of institutions in both civil society and the public sector. These new institutions should bring about greater accountability for AI harms and diffuse the concentration of power and expertise in the hands of the tech companies and venture capitalists driving much of AI development and decision making today. Philanthropy has done this before in the domain of technology, in areas such as open-source ecosystems and digital public infrastructure (White 2023). A key outcome for such efforts is to ensure that new bodies and institutions have a harmonious relationship with existing authorities and serve to both assist them (for example, with nimble research or rapid-response trials, or with diverse perspectives or community engagement) and hold them accountable to the public interest.

Philanthropy can help leaders and authorities incentivize collaboration and mechanisms that lead to world-class progress. The way forward will require both government and industry to establish well-designed accountability systems with appropriate guardrails and checks and balances to prevent serious harm, establish liability, create public alternatives and provide remedies and redress. Philanthropy's substantive engagement, done well, should accelerate a more equitable technology ecosystem.

16 See https://new.nsf.gov/focus-areas/artificial-intelligence/nairr.

Racing Against the Race to the Bottom

The first public report of the UN Secretary-General's high-level advisory body on AI, *Governing AI for Humanity*, notes that while there are numerous existing governance models that offer inspiration for new intergovernmental bodies designed to govern AI, there is no obvious single best choice (Advisory Body on Artificial Intelligence 2023, 13). The report identifies seven "AI Governance Functions" that any future institutions will need (ibid., 16). The member states of the United Nations are currently negotiating a Global Digital Compact, described in a recently released "zero draft," which has among its objectives "Govern[ing] emerging technologies, including Artificial Intelligence, for humanity" and whose adoption is expected in September 2024 (United Nations 2024, 1). The zero draft also calls for the creation of an "International Scientific Panel on AI" (ibid., 10) that appears to be at least partly modelled on the Intergovernmental Panel on Climate Change, as well as a "Global Fund for AI and Emerging Technologies for Sustainable Development" (ibid., 11).

Through these documents, it is evident that the United Nations sees a need for the rapid launch of institutions capable of monitoring risks and opportunities, rendering governance and technical standards interoperable, and harnessing AI for the public interest, all this with a focus on the United Nations' SDGs. In email communications with UN officials involved in this effort, we learned that the intent is for this work to be initially supported by a small team that will transition into a full-fledged UN AI Office by the end of 2025.

Under normal circumstances, this timeline for global action might appear to be reasonable. However, for those of us concerned with the urgency of the pacing problem and the fact that each day of delay in establishing a governing body perpetuates a "race to the bottom" in rapidly deploying AI systems, it seems woefully inadequate.

A particularly striking reminder of the "race to the bottom" was a recent study that found that a version of Stable Diffusion, one of the most popular open-source AI image-generating tools, had been trained with thousands of CSAM images, which would presumably allow it to be effective at generating AI-generated child sexual abuse material (AIG-CSAM) (Thiel 2023). Another recent study found that AI "undressing" tools, which can turn an image of a clothed person (although many of the tools only work on women) into NCII are gaining popularity, with at least 34 such tools widely available online for public use (Lakatos 2023). The growing availability of open-source AI image-generation tools has been cited as a key factor in enabling the creation of these undressing tools. As one example of the real-life harms these tools can enable, at a high school in New Jersey, dozens of pornographic images of female students were created, with devastating effects on the children involved (J. Downes 2023).

If only a few jurisdictions worldwide take action against NCII generators and AIG-CSAM, it will not be effective. AI developers in unregulated locales will simply continue to build these tools, unless they are held accountable (Kalia, Ponnezhath and Arunasalam 2024). If the makers of popular tools are careless in their production and those tools are trained on illegal CSAM images, they should be held accountable as well. The proliferation of AI systems that either intentionally or recklessly provide people with tools that can be wielded with devastating consequences should be a crime. However, it is not yet clear today whether the makers of these tools will be, or even can be, held liable under existing law.

The Big Bet: The IIAII

Law makers are scrambling to address a multitude of urgent AI harms such as those discussed above in numerous jurisdictions, but without a centralized, international body to coordinate and support these efforts, they are likely to take far too long to be effective against these already occurring harms and those on the immediate horizon (Nilsson 2017).

The project of governing AI needs to unfold in dialogue with existing frameworks, such as the United Nations Educational, Scientific and Cultural Organization's global agreement on AI ethics and the Group of Seven's Hiroshima AI Process. While these frameworks are acknowledged in the UN interim report, there is a need to clarify how any new institutional function would interact with

them. This process requires the active participation of numerous intergovernmental organizations. Through close communication and collaboration, a new organization will need to work quickly to "harmonize standards, safety, and risk management frameworks" (Advisory Body on Artificial Intelligence 2023, 16). Thus, the opportunity for the philanthropic big bet. Leveraging the unique power of strategic philanthropists discussed above, there is an opportunity to work alongside the creation of the emerging UN AI Office, in a way that follows the principles discussed above, as well as those already enumerated in the UN AI Advisory Body's interim report.

To move as quickly as possible against the race to the bottom and toward maximum common benefit, we propose that one or more philanthropists make a gift in the tens of millions of dollars to fund the first three years of operation of the IIAII, so that even if the creation of the United Nations' AI Office is delayed significantly, this body can operate in its absence. At the end of this initial three-year period, or hopefully sooner, the IIAII and its assets could be placed under UN ownership.

Within this proposed body's name, the initial word, *interim*, is important — it signals humility through the understanding that such a body must have a democratic mandate to be successful in the long run. The final term, *institution*, is also significant in its flexibility — it acknowledges that the entity could evolve in the future into an agency or body of another institution or an independent organization — again in preparation for possible unforeseen challenges within the UN system. It would bring together the highest level of technical and policy expertise in the service of rapid-response regulatory development. A key goal during this period would be to facilitate strong legal standards and enforcement mechanisms to bring AI under democratic control. Tasks that the IIAII could undertake would include three key areas, as follows.

Rapid-Response Harmonized Policy Development Support

The IIAII's first priority would be in providing rapid-response support to regulatory or legislative bodies anywhere in the world working to develop AI regulation. By assisting national governments in the development of AI policy, the institution can get a head start on the process of harmonizing — to the greatest extent possible — AI regulations around the globe. This support could include legal, technical, translation, communications and even legislative strategy support — everything that regulators need to ensure that their AI laws are robust and iterative, passed as quickly as possible, and harmonized with other jurisdictions (and eventually able to integrate with an international legal regime), at the state, national or international level. Even in a place such as California, we have seen firsthand that legislators lack access to the combination of skilled technologists and policy experts who understand the legal and multidisciplinary issues at hand. We have seen this gap lead to confusion, overreach or duplication of effort that unnecessarily hinders democratic oversight via the regulatory process.

Some examples of ways in which the IIAII could provide support include:

- **Prototyping a global licensing and registration standard for AI systems.** The prototyping work could define all the details of how such a standard would be implemented and then make it accessible, as soon as possible, so that when such a system becomes legally binding in certain jurisdictions, it is already available and tested.
- **Developing best-practice monitoring mechanisms to detect AI harms and risks.** The use of these monitoring mechanisms could then be mandated by regulation. Their functions could include detecting AI-powered misinformation campaigns and influence operations taking place on social media and monitoring the proliferation of unlicensed AI systems that could pose harm; monitoring of the "dark web" to find AI systems that can be used to produce NCII or CSAM; and building tools to detect content from AI systems that do not use required watermarking standards.
- **Building a global version of the planned NAIRR.** This task would entail both acquiring technology capacity directly in the public sector and requiring that AI companies reserve a certain amount of their capacity for use by vetted researchers at universities and research institutions. This parallel approach would ensure that the IIAII builds toward a "public option" that democratizes data, compute and distribution of access to AI resources without bankrupting public sector institutions.

→ **Providing technical support for the development of key requirements, standards and best practices for AI developers.** Fulfilling this task would ensure that AI developers are not left uncertain about the expectations for how they can develop AI systems responsibly, and might lead to the creation of the following:

- guidelines for assessing the human rights impacts of AI systems, including but not limited to the fairness of, or potential discrimination caused by, these systems, or the ways in which these systems might violate privacy;
- guidelines for how to transparently disclose what training data was used to produce AI systems and how to make sure that this training data was ethically and legally sourced;
- guidelines for how to conduct adversarial testing or red teaming of AI systems before they are released to the public; and
- guidelines for understanding the potential harms of AI systems across different applications.

Support Collaborative Standards Development

We are currently seeing a proliferation of efforts to establish standards around AI, with different national and international bodies operating independently, for example, the US National Institute for Standards and Technology, the European Union's CEN-CENELEC,[17] the International Organization for Standardization and the Institute of Electrical and Electronics Engineers. A technically sophisticated coordinating body is needed to facilitate collaboration between these institutions to accelerate the development of AI standards, such as a maximally indelible watermarking technique (a way for generative AI systems to mark the content they produce with a unique identifiable signature that is difficult to remove, and that can be detected and displayed wherever the content is viewed). Standards such as this — or common language to set up requirements for industry best practices — could be developed in-house, or through coordination between existing groups working on similar standards. In the case of watermarking, this could mean bridging the gap between the Adobe-led Coalition for Content Provenance and Authenticity[18] and efforts already in place in China.[19]

Support Coordinated Public-Interest AI Development

In collaboration with universities and research centres around the world, the IIAII could support the coordination of world-class research and help to productively direct the allocation of AI resources to solve scientific problems in areas of public interest, such as climate change, misinformation, public health, energy, agriculture and many other fields. The United Nations' SDGs present a strong set of guiding goals that could be used to focus these efforts.

The IIAII could also support increased access to AI education by offering public educational resources, including online courses taught by IIAII staff members working on all of the items above and designed to democratize access to AI knowledge and spread its benefits widely.

Succession Planning

The final duty of the IIAII is long-term succession planning for the institution itself. This would involve regular documentation of successes and failures in its efforts at the above goals, such that it could make recommendations based on its own experience of what its successor institution(s) should look like. It should also strive to operate flexibly to prepare for that transition, using UN principles and working practices to guide its efforts and smooth a possible transition into the UN system.

While it may eventually make sense to have these types of activities in different institutional homes, there is value in having a broad set of such activities under one roof at the outset, as it has the potential to spark synergies among a multidisciplinary group of people who would otherwise have to fly across the world to connect and exchange ideas face to face.

17 In full, the European Committee for Standardization and the European Electrotechnical Committee for Standardization.

18 See https://c2pa.org/.

19 See www.insideglobaltech.com/2023/08/30/labeling-of-ai-generated-content-new-guidelines-released-in-china/.

It is our contention that the combined cost of these activities — including the convening, travel and maintenance of a high-level team of 50 to 60 experts working full-time in the same location for three years — should be in the tens of millions of dollars, easily funded by the philanthropic community already backing AI initiatives in this vein.

Conclusion

Philanthropic capital is uniquely positioned to make this investment in the creation of the envisioned IIAII, given its ability to make fast decisions and take bold risks in the public interest. This moment represents a unique leverage point, where a vacuum has emerged and technological advances are far outpacing regulatory action.

The recent successful passage of the EU AI Act (Bertuzzi 2024) is both a sign of good things to come and an indicator of the inadequately slow pace of government action, even in the most advanced democracy that has managed to pass meaningful AI regulation. Some of the provisions of the EU AI Act will come into force after six months, but others will require between one and three years to take effect. Given the relative slowness of the US Congress to regulate social media technology and online privacy, it is hard to imagine that legislation not yet even introduced in the US Congress will come into force before the EU AI Act (Pearlstein 2023).

The longer we wait to regulate AI, the more people will be harmed by either careless or malicious use of AI tools. As discussed in this paper's introduction, some of the often irreparable harms caused by unregulated AI systems include discrimination in lending, housing, employment, health care, and many other areas; the production of interactive misinformation, disinformation and malinformation; election interference; and the production of NCII. For every month that we wait to regulate AI, more and more powerful unsecured AI tools proliferate, and it will be very difficult to ever get these tools out of circulation. Therefore, propelling regulation forward as quickly as possible is a critically time-sensitive effort.

Another reason that philanthropy may be the best, and perhaps only, option to fund such an endeavour is the competitive dynamics inherent to the relationship between adversarial nations. The EU AI Act notably created exemptions for national security and police use of many types of otherwise prohibited AI systems (Nolan, Maryam and Kleinman 2024). As with climate change, regulation and nuclear non-proliferation agreements, many, if not most, countries do not want to be the first to make bold commitments to self-restraint without knowing that other nations will follow suit. This "regulation dilemma," arguably more difficult to overcome than the prisoner's dilemma (Han et al. 2021), is further evidence of the need for a global coordinating body to step in as soon as possible to drive forward global collaboration and progress on AI regulation.

Ted Turner's billion-dollar donation to the United Nations is still remembered a quarter-century later. The creation of the IIAII is an opportunity to make an outsized impact at a critical inflection point for AI that will deliver broad societal benefits well into the next century.

Acknowledgements

The authors would like to thank the research team from the University of California, Berkeley, that supported this work: Owen Doyle, Milad Brown, Anish Ganga, Nancy Hu, Ashley Chan, Ruiyi Chen, Maddy Cooper, Daniel Jang, Parth Shinde, and Vinaya Sivakumar. This work also benefited significantly from feedback from a senior United Nations official. For David Evan Harris, additional thanks are due to the Geneva Centre for Philanthropy at the University of Geneva, CIGI, the International Computer Science Institute, and the John S. and James L. Knight Foundation for supporting the research that made this work possible. For Anamitra Deb, additional thanks go to Omidyar Network, where his position has allowed him to participate in such timely, responsible technology discussions.

Works Cited

Advisory Body on Artificial Intelligence. 2023. *Interim Report: Governing AI for Humanity*. New York, NY: United Nations. www.un.org/ai-advisory-body.

Agüera y Arcas, Blaise. 2023. "Fears about AI's existential risk are overdone, says a group of experts." *The Economist*, July 21. www.economist.com/by-invitation/2023/07/21/fears-about-ais-existential-risk-are-overdone-says-a-group-of-experts.

Akselrod, Olga and Cody Venzke. 2023. "How Artificial Intelligence Might Prevent You From Getting Hired." *American Civil Liberties Union* (blog), August 23. www.aclu.org/news/racial-justice/how-artificial-intelligence-might-prevent-you-from-getting-hired.

Backman, Isabella. 2023. "Eliminating Racial Bias in Health Care AI: Expert Panel Offers Guidelines." Yale School of Medicine, December 21. https://medicine.yale.edu/news-article/eliminating-racial-bias-in-health-care-ai-expert-panel-offers-guidelines/.

Banco, Erin, Ashleigh Furlong and Lennart Pfahler. 2022. "How Bill Gates and partners used their clout to control the global Covid response — with little oversight." Politico, September 14. www.politico.com/news/2022/09/14/global-covid-pandemic-response-bill-gates-partners-00053969.

Belfield, Haydn. 2020. "Activism by the AI Community: Analysing Recent Achievements and Future Prospects." In *AIES '20: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 15–21. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3375627.3375814.

Bertuzzi, Luca. 2024. "EU countries give crucial nod to first-of-a-kind Artificial Intelligence law." Euractiv, February 2. www.euractiv.com/section/artificial-intelligence/news/eu-countries-give-crucial-nod-to-first-of-a-kind-artificial-intelligence-law/.

Betsill, Michele M., Ashley Enrici, Elodie Le Cornu and Rebecca L. Gruby. 2022. "Philanthropic foundations as agents of environmental governance: a research agenda." *Environmental Politics* 31 (4): 684–705. https://doi.org/10.1080/09644016.2021.1955494.

Bosire, Lydiah Kemunto. 2020. "How to Make Society's Risk Capital Riskier." Stanford Social Innovation Review, July 29. https://ssir.org/articles/entry/how_to_make_societys_risk_capital_riskier.

Bradford, Anu. 2020. *The Brussels Effect: How the European Union Rules the World*. New York, NY: Oxford University Press.

Buck, Stuart. 2020. "How Philanthropy Could Embrace More High-Risk, High-Reward Projects." *Inside Philanthropy*, September 9. www.insidephilanthropy.com/home/2020/9/9/how-philanthropy-could-embrace-more-high-risk-high-reward-projects.

Carbonaro, Giulia. 2023. "How is the World Health Organization funded, and why does it rely so much on Bill Gates?" Euronews, March 2. www.euronews.com/health/2023/02/03/how-is-the-world-health-organization-funded-and-why-does-it-rely-so-much-on-bill-gates.

Chan, Kelvin. 2023. "Europe agreed on world-leading AI rules. How do they work and will they affect people everywhere?" Associated Press, December 11. https://apnews.com/article/eu-ai-act-artificial-intelligence-regulation-0283a10a891a24703068edcae3d60deb.

Chatterjee, Mohar and Gian Volpicelli. 2023. "France bets big on open-source AI." Politico, August 4. www.politico.eu/article/open-source-artificial-intelligence-france-bets-big/.

Chee, Foo Yun. 2023. "Exclusive: AI rules 'cannot be bargained', EU's Breton says after OpenAI CEO threat." Reuters, May 25. www.reuters.com/technology/eus-breton-slams-openai-ceos-comments-blocs-draft-ai-rules-2023-05-25/.

CNN. 1997. "Ted Turner donates $1 billion to 'U.N. causes.'" CNN.com, September 19. http://edition.cnn.com/US/9709/18/turner.gift/.

Covington & Burling LLP. 2023. "U.S. Artificial Intelligence Policy: Legislative and Regulatory Developments." Alert, October 20. www.cov.com/en/news-and-insights/insights/2023/10/us-artificial-intelligence-policy-legislative-and-regulatory-developments.

Crossette, Barbara. 1997. "U.S. Effort to Cut Its Dues Dies at an Angry U.N." *The New York Times*, December 19. www.nytimes.com/1997/12/19/world/us-effort-to-cut-its-dues-dies-at-an-angry-un.html.

De Tena, Carlos Luca, Diego Rubio, Oscar Jonsson, Dario Garcia De Viedma, Carlos Lastra Anadon, Irene Menendez, Carl Benedikt Frey and Christina J. Colclough. 2023. "European Tech Insights 2023." Madrid, Spain: IE University Center for the Governance of Change. www.ie.edu/cgc/research/european-tech-insights/#authors.

Downes, Jessica. 2023. "AI-Generated CSAM: The Call for Proactive Action." *CHILD USA* (blog), November 21. https://childusa.org/ai-generated-csam-the-call-for-proactive-action/.

Downes, Larry. 2009. *The Laws of Disruption: Harnessing the New Forces That Govern Life and Business in the Digital Age.* 1st ed. New York, NY: Basic Books.

Field, Hayden. 2023. "OpenAI's Sam Altman reverses threat to cease European operations." CNBC, May 26. www.cnbc.com/2023/05/26/openai-ceo-sam-altman-reverses-threat-to-cease-european-operations.html.

Ford Foundation. 2023. "Philanthropies launch new initiative to ensure AI advances the public interest." News release, November 1. www.fordfoundation.org/news-and-stories/news-and-press/news/philanthropies-launch-new-initiative-to-ensure-ai-advances-the-public-interest/.

Grace, Katja, Harlan Stewart, Julia Fabienne Sandkühler, Stephen Thomas, Ben Weinstein-Raun and Jan Brauner. 2024. "Thousands of AI Authors on the Future of AI." arXiv. January. https://arxiv.org/abs/2401.02843.

Han, The Anh, Francisco C. Santos, Luís Moniz Pereira and Tom Lenaerts. 2021. "A Regulation Dilemma in Artificial Intelligence Development." In *Proceedings of the ALIFE 2021: The 2021 Conference on Artificial Life*. Cambridge, MA: MIT Press. https://doi.org/10.1162/isal_a_00385.

Harris, David Evan. 2023a. "AI is already causing unintended harm. What happens when it falls into the wrong hands?" *The Guardian*, June 16. www.theguardian.com/commentisfree/2023/jun/16/ai-new-laws-powerful-open-source-tools-meta.

———. 2023b. "Biden's Executive Order on AI: Bravo, Now Back to Work." Opinion, Centre for International Governance Innovation, November 8. www.cigionline.org/articles/bidens-executive-order-on-ai-bravo-now-back-to-work/.

———. 2023c. "How to Regulate Unsecured 'Open-Source' AI: No Exemptions." Tech Policy Press, December 4. https://techpolicy.press/how-to-regulate-unsecured-opensource-ai-no-exemptions.

———. 2023d. "Europe has made a great leap forward in regulating AI. Now the rest of the world must step up." *The Guardian*, December 13. www.theguardian.com/commentisfree/2023/dec/13/europe-regulating-ai-artificial-intelligence-threat.

Heath, Ryan. 2023. "China races ahead of U.S. on AI regulation." Axios, May 8. www.axios.com/2023/05/08/china-ai-regulation-race.

Hendrycks, Dan, Mantas Mazeika and Thomas Woodside. 2023. "An Overview of Catastrophic AI Risks." arXiv. http://arxiv.org/abs/2306.12001.

Ho, Lewis, Joslyn Barnhart, Robert Trager, Yoshua Bengio, Miles Brundage, Allison Carnegie, Rumman Chowdhury, et al. 2023. "International Institutions for Advanced AI." arXiv. https://doi.org/10.48550/arXiv.2307.04699.

Hu, Krystal. 2023. "ChatGPT sets record for fastest-growing user base — analyst note." Reuters, February 2. www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/.

Hutchinson, Christophe Samuel. 2022. "Potential abuses of dominance by big tech through their use of Big Data and AI." *Journal of Antitrust Enforcement* 10 (3): 443–68. https://doi.org/10.1093/jaenfo/jnac004.

Johnson, Khari. 2023. "The US Has Failed to Pass AI Regulation. New York City Is Stepping Up." *Wired*, October 19. www.wired.com/story/us-failed-to-pass-ai-regulation-new-york-city-stepping-up/.

Kalia, Shubham, Maria Ponnezhath and Samrhitha Arunasalam. 2024. "Musk seeks Tesla shareholder vote on moving incorporation to Texas." Reuters, February 1. www.reuters.com/business/autos-transportation/tesla-take-shareholder-vote-shifting-incorporation-texas-musk-says-2024-02-01/.

Kang, Cecilia. 2023. "2 Senators Propose Bipartisan Framework for A.I. Laws." *The New York Times*, September 7. www.nytimes.com/2023/09/07/technology/artificial-intelligence-framework-senate.html.

Knapp, Alex. 2023. "New XPrize Will Award $101 Million To Innovators Who Can Reverse Aging." *Forbes*, November 29. www.forbes.com/sites/alexknapp/2023/11/29/new-xprize-will-award-101-million-to-innovators-who-can-reverse-aging/.

Kubzansky, Mike. 2024. "Addressing AI's present and future: a false trichotomy?" Context, May 7. www.context.news/ai/opinion/addressing-ais-present-and-future-a-false-trichotomy.

Kuokkanen, Tuomas and Yulia Yamineva. 2013. "Regulating Geoengineering in International Environmental Law." *Carbon & Climate Law Review* 7 (3): 161–67. www.jstor.org/stable/24323970.

Lakatos, Santiago. 2023. *A Revealing Picture — AI-Generated "Undressing" Images Move from Niche Pornography Discussion Forums to a Scaled and Monetized Online Business.* December. New York, NY: Graphika. https://public-assets.graphika.com/reports/graphika-report-a-revealing-picture.pdf.

Larson, Jeff, Surya Mattu, Lauren Kirchner and Julia Angwin. 2016. "How We Analyzed the COMPAS Recidivism Algorithm." ProPublica, May 23. www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

Law, Tara. 2023. "Laurene Powell Jobs: Let's Tackle Climate Change 'Like a Speed Boat' Before The Window Closes." *Time*, April 25. https://time.com/collection/earth-awards-2023/6274438/laurene-powell-jobs-climate-philanthropy/.

Leprince-Ringuet, Daphné. 2024. "Mistral's deal with Microsoft is causing controversy." Sifted, February 28. https://sifted.eu/articles/mistral-microsoft-deal-controversy/.

Malik, Hanna Maria, Mika Viljanen, Nea Lepinkäinen and Anne Alvesalo-Kuusi. 2022. "Dynamics of Social Harms in an Algorithmic Context." *International Journal for Crime, Justice and Social Democracy* 11 (1): 182–95. https://doi.org/10.5204/ijcjsd.2141.

Marchant, Gary E. 2011. "The Growing Gap Between Emerging Technologies and the Law." In *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem*, edited by Gary E. Marchant, Braden R. Allenby and Joseph R. Herkert, 19–33. The International Library of Ethics, Law and Technology, Vol. 7. Dordrecht, The Netherlands: Springer.

McCracken, Harry. 2023. "Microsoft's Satya Nadella is winning Big Tech's AI war. Here's how." *Fast Company*, August 21. www.fastcompany.com/90931084/satya-nadella-microsoft-ai-frontrunner.

McGuigan, Elizabeth and David Bass. 2022. "Policy Philanthropy and Its Key Role in Civil Society." Policy Primer, November. Washington, DC: Philanthropy Roundtable. www.philanthropyroundtable.org/wp-content/uploads/2022/11/Policy-Philanthropy-and-Its-Key-Role-in-Civil-Society.pdf.

McIlwain, Charlton. 2020. "AI Has Exacerbated Racial Bias in Housing. Could It Help Eliminate It Instead?" *MIT Technology Review*, October 20. www.technologyreview.com/2020/10/20/1009452/ai-has-exacerbated-racial-bias-in-housing-could-it-help-eliminate-it-instead/.

Moore, Martin and Damian Tambini, editors. 2022. *Regulating Big Tech: Policy Responses to Digital Dominance*. New York, NY: Oxford University Press.

Morena, Edouard. 2023. "The Spirit of Climate Philanthropy." In *The Routledge International Handbook of Critical Philanthropy and Humanitarianism*, edited by Katharyne Mitchell and Polly Pallister-Wilkins, 248–58. Abingdon, UK: Routledge.

NBC News. 2005. "Gates' foundation donates $750 million." NBC News, January 25. www.nbcnews.com/health/health-news/gates-foundation-donates-750-million-flna1C9440090.

Nichols, Michelle. 2023. "UN chief backs idea of global AI watchdog like nuclear agency." Reuters, June 12. www.reuters.com/technology/un-chief-backs-idea-global-ai-watchdog-like-nuclear-agency-2023-06-12/.

Nilsson, Adriana. 2017. "Making norms to tackle global challenges: The role of Intergovernmental Organisations." *Research Policy* 46 (1): 171–81. https://doi.org/10.1016/j.respol.2016.09.012.

Nolan, David, Hajira Maryam and Michael Kleinman. 2024. "The Urgent but Difficult Task of Regulating Artificial Intelligence." Amnesty International, January 16. www.amnesty.org/en/latest/campaigns/2024/01/the-urgent-but-difficult-task-of-regulating-artificial-intelligence/.

Omidyar Network. 2022. "Our Vision for a Responsible Tech Future." Omidyar Network, September 20. https://omidyar.com/responsible-tech-future/.

O'Reilly, Thomas. 2023. "EU's AI Act Slammed As Innovation Killer." *The European Conservative*, December 14. https://europeanconservative.com/articles/news/eus-ai-act-slammed-as-innovation-killer/.

Orth, Taylor. 2023. "Americans are divided on AI's societal impact, but most support government regulation." YouGov, May 25. https://today.yougov.com/politics/articles/45747-americans-are-divided-artificial-intelligence-poll.

Partnership on AI. 2020. *2019 Annual Report: Building a Connected Community for Responsible AI*. San Francisco, CA: Partnership on AI. http://partnershiponai.org/wp-content/uploads/2021/08/PAI-2019-Annual-Report.pdf.

Pearlstein, Steven. 2023. "Here's the inside story of how Congress failed to rein in Big Tech." *Washington Post*, July 6. www.washingtonpost.com/opinions/2023/07/06/congress-facebook-google-amazon-apple-regulation-failure/.

Robinson, Jemima and Matthew Edwards. 2024. "Fraudsters target the elderly: Behavioural evidence from randomised controlled scam-baiting experiments." *Security Journal*, January. https://doi.org/10.1057/s41284-023-00410-4.

Rubinson, Paul. 2021. "Philanthropy, Nuclear Nonproliferation, and Threat Reduction." February. Washington, DC: Urban Institute. www.urban.org/sites/default/files/2021/02/05/philanthropy_nuclear_nonproliferation_and_threat_reduction.pdf.

Russell, Stuart, Karine Perset and Marko Grobelnik. 2023. "Updates to the OECD's definition of an AI system explained." OECD.AI Policy Observatory, November 29. https://oecd.ai/en/wonk/ai-system-definition-update.

Satter, Raphael. 2024. "Microsoft says it caught hackers from China, Russia and Iran using its AI tools." Reuters, February 14. www.reuters.com/technology/cybersecurity/microsoft-says-it-caught-hackers-china-russia-iran-using-its-ai-tools-2024-02-14/.

Schwartz, Debra. 2024. "Finding Hope in Catalytic Capital and its Champions." *MacArthur Foundation* (blog), February 1. www.macfound.org/press/perspectives/finding-hope-in-catalytic-capital-and-its-champions.

Seger, Elizabeth, Noemi Dreksler, Richard Moulange, Emily Dardaman, Jonas Schuett, K. Wei, Christoph Winter, et al. 2023. "Open-Sourcing Highly Capable Foundation Models: An evaluation of risks, benefits, and alternative methods for pursuing open-source objectives." LawAI Working Paper No. 2-2023. Centre for the Governance of AI, October. https://doi.org/10.2139/ssrn.4596436.

Sheehan, Matt. 2023. "China's AI Regulations and How They Get Made." Carnegie Endowment for International Peace, July 10. https://carnegieendowment.org/2023/07/10/china-s-ai-regulations-and-how-they-get-made-pub-90117.

Shubinski, Barbara. 2022. "'A Roomful of Brains': Early Advances in Computer Science and Artificial Intelligence." *RE:source*, January 6. Sleepy Hollow, NY: Rockefeller Archive Center. https://resource.rockarch.org/story/a-roomful-of-brains-early-advances-in-computer-science-and-artificial-intelligence/.

Sitaraman, Ganesh and Natalie Foster. 2024. "How States Can Keep Big Tech from Dominating AI." Politico, March 13. www.politico.com/news/magazine/2024/03/13/states-big-tech-ai-00146338.

Sklair, Jessica and Paul Gilbert. 2022. "Giving as 'De-Risking': Philanthropy, Impact Investment and the Pandemic Response." *Public Anthropologist* 4 (1): 51–77. https://doi.org/10.1163/25891715-bja10033.

Slaughter, Anne-Marie and Darren Walker. 2021. "Afterword." In *Power to the Public: The Promise of Public Interest Technology*, by Tara Dawson McGuinness and Hana Schank, 141–52. Princeton, NJ: Princeton University Press.

Sparke, Matthew and Orly Levy. 2022. "Competing Responses to Global Inequalities in Access to COVID Vaccines: Vaccine Diplomacy and Vaccine Charity Versus Vaccine Liberty." *Clinical Infectious Diseases* 75 (Supplement 1): S86–S92. https://doi.org/10.1093/cid/ciac361.

Stanford Internet Observatory. 2023. "Trust & Safety Research Conference showcases leading research in preventing online harms." News, October 25. https://cyber.fsi.stanford.edu/io/news/trust-safety-research-conference-showcases-leading-research-preventing-online-harms.

Suleyman, Mustafa and Eric Schmidt. 2023. "We need an AI equivalent of the IPCC." *Financial Times*, October 19. www.ft.com/content/d84e91d0-ac74-4946-a21f-5f82eb4f1d2d.

Thakur, Dhanaraj and DeVan Hankerson Madrigal. 2021. "Facts and their Discontents: A Research Agenda for Online Disinformation, Race, and Gender." Center for Democracy & Technology, February 11. https://cdt.org/insights/facts-and-their-discontents-a-research-agenda-for-online-disinformation-race-and-gender/.

The New York Times. 1997. "A Start on U.N. Reform." March 18. www.nytimes.com/1997/03/18/opinion/a-start-on-un-reform.html.

The Rockefeller Foundation. 2021. "Catalytic Capital Consortium Announces $1.2 Million to Leading Impact Investing Networks." Press release, July 28. www.rockefellerfoundation.org/news/catalytic-capital-consortium-announces-1-2-million-to-leading-impact-investing-networks/.

———. 2022. "The Rockefeller Foundation Announces Nearly $500K to Combat AI Bias and Discrimination." Press release, March 28. www.rockefellerfoundation.org/news/the-rockefeller-foundation-announces-nearly-500k-to-combat-ai-bias-and-discrimination/.

Thiel, David. 2023. "Identifying and Eliminating CSAM in Generative ML Training Data and Models." Stanford Digital Repository, December 20. https://purl.stanford.edu/kh752sm9123.

Thrower, Sharece. 2021. "What is an executive order, and why don't presidents use them all the time?" The Conversation, January 26. http://theconversation.com/what-is-an-executive-order-and-why-dont-presidents-use-them-all-the-time-150896.

Toma, Alexandra. 2022. *The Changing Landscape of Nuclear Security Philanthropy: Risks and Opportunities in the Current Moment*. June. San Francisco, CA: Founders Pledge Patient Philanthropy Fund. https://static1.squarespace.com/static/62435b2d773bcf0ad9cbb672/t/62cee05cb38a4b2487a533e0/1657725020334/The+Changing+Landscape+of+Nuclear+Security+Philanthropy.pdf.

Trager, Robert F., Ben Harack, Anka Reuel, Allison Carnegie, Lennart Heim, Lewis Ho, Sarah Kreps, et al. 2023. "International governance of civilian AI: A jurisdictional certification approach." LawAI Working Paper No.3-2023, August. https://doi.org/10.2139/ssrn.4579899.

Turner, Craig. 1997. "Ted Turner Makes $1-Billion Pledge to U.N. Programs." *Los Angeles Times*, September 19. www.latimes.com/archives/la-xpm-1997-sep-19-mn-34095-story.html.

Turner, Seth. 2016. "Framers Knew That Slow and Steady Wins the Race." *Congressional Management Foundation* (blog), September 16. Washington, DC: Congressional Management Foundation. www.congressfoundation.org/news/blog/1183.

United Nations. 2006. "Ted Turner's United Nations Foundation Delivers $1 Billion to UN Causes." Press release, October 11. New York, NY: UN Department of Public Information. https://press.un.org/en/2006/dev2594.doc.htm.

———. 2024. "Global Digital Compact: zero draft." April 1. New York, NY: Office of the UN Secretary-General's Envoy on Technology. www.un.org/techenvoy/sites/www.un.org.techenvoy/files/Global_Digital_Compact_Zero_Draft.pdf.

US Department of State. 1997. "U.S. Agenda for UN Reform." Bureaus of International Organization Affairs and Public Affairs, US Department of State, November 29. https://1997-2001.state.gov/issues/fs_us_agend_un_reform_0997.html.

Wanat, Zosia. 2023. "'EU's AI act could kill our company,' says Mistral's Cédric O." Sifted, October 27. https://sifted.eu/articles/eu-ai-act-kill-mistral-cedric-o/.

White, Joseph. 2023. "Musk: 'AI stresses me out.'" Reuters, March 1. www.reuters.com/business/autos-transportation/musk-ai-stresses-me-out-2023-03-02/.

Wong, Scott, Frank Thorp V, Ryan Nobles and Liz Brown-Kaiser. 2023. "Elon Musk warns of 'civilizational risk' posed by AI in meeting with tech CEOs and senators." NBC News, September 13. www.nbcnews.com/politics/congress/big-tech-ceos-ai-meeting-senators-musk-zuckerberg-rcna104738.

Zakrzewski, Cat, Cristiano Lima-Strong and David DiMolfetta. 2023. "Tech leaders including Musk, Zuckerberg call for government action on AI." *The Washington Post,* September 14. www.washingtonpost.com/technology/2023/09/13/senate-ai-hearing-musk-zuckerburg-schumer/.

Zhao, Christina. 2018. "'Black Mirror' in China? 1.4 Billion Citizens to Be Monitored Through Social Credit System." *Newsweek,* May 1. www.newsweek.com/china-social-credit-system-906865.

Centre for International
Governance Innovation

67 Erb Street West
Waterloo, ON, Canada N2L 6C2
www.cigionline.org