global
POLICY

GOVERNING ARTIFICIAL INTELLIGENCE

Edited by LAURA MAHRENBACH

Durham University
School of Government
and International Affairs

Emerging Global Governance
LEADING REASEARCH AND KNOWLEDGE MOBILIZATION

Cover Image: Foto von Jeremy Waterhouse

Mahrenbach, E. Ed. (2023) Governing Artificial Intelligence. Durham: Global Policy.

CONTENTS

Introduction - Aligning Global Threats and Opportunities via AI Governance: A Commentary Series

Laura C. Mahrenbach

Artificial intelligence (AI) is here to stay. Historically, the complicated technology and high costs affiliated with these technologies have meant that AI was depicted as something from the realm of science fiction – or at least something separate from everyday life. Yet AI is increasingly crucial to daily decision-making and activities, whether we are aware of it or not.

For example, GoogleMaps uses AI to improve road safety on its proposed routes and to generate accurate maps for pedestrians and bicyclists. Similarly, business applications of AI are numerous, including customer service chatbots, e-mail spam filters, and tools for tailoring suggestions to user preferences. Meanwhile, government buildings around the world use AI-based facial recognition for security purposes and doctors use machine learning to diagnose diseases and personalize treatment.

In this context, two narratives tend to shape our understandings of how AI may impact social relations. One narrative focuses on the opportunities offered by AI and related digital technologies. As evident above, private actors are developing commercial uses for AI technology which improve patient care, enhance global communication and keep your floors clean. Governments see opportunities as well. Meetings like the International Telecommunication Union's annual AI for Good Summit generate ideas about how AI can be used to achieve socially beneficial goals and facilitate the private-public connections necessary to enact these ideas. Similarly, governments around the world are also forging ahead alone, with the Global AI Index illustrating both the worldwide commitment to funding AI-enabled innovation and the disparity among nations in their capacity to do so.

Another narrative exists as well: AI as a threat. The threat narrative tends to be interlinked with an inter-state, geopolitical-strategic competition approach to technological development. While the dual-use capacity of AI provides ample incentives for private actors to push innovations in AI, competition to secure a limited supply of top AI research talent and to shape rules governing acceptable AI use ensures growing government involvement in setting the tone – and funding – AI innovations. Part of the threat posed by AI refers to its military uses. These have been (violently) on display following Russia's invasion of the Ukraine, raising concerns that, in using autonomous weapons, we inhibit our capacity to protect human lives without using AI technology. Yet threats are evident beyond the battlefield as well. The EU's proposed AI Act, for instance, discusses the threats posed by AI to employment, access to public services, privacy, human rights and judicial procedures across Europe, among others.

Effective governance is the linchpin necessary to address and align solutions to these two narratives. As ChatGPT noted (when asked by me), "AI has the potential to bring significant benefits to society, but it is crucial to manage its implementation and ensure that it is used in an ethical and responsible way" (see Figure 1). This may not be easy. One issue

involves convincing governments to agree to rules which could put them at a geopolitical, economic or military disadvantage. Another involves the domestic consequences of regulations for private and societal actors, which can hinder political agreement despite publics largely remaining silent on these topics. A related issue involves identifying which definition(s) of AI are relevant for achieving intended regulatory goals. And then there is the question of where to regulate AI.

Figure 1. ChatGPT on how AI affects society (question posed on February 27, 2023)

LA How does AI affect society?

Artificial intelligence (AI) has the potential to significantly impact society in various ways, including:

1. Automation: AI can automate certain tasks that were previously performed by humans, leading to increased efficiency and productivity. However, this could also result in job loss and economic disruption.
2. Personalization: AI can help businesses personalize their services and products to meet individual needs, improving customer satisfaction and loyalty.
3. Healthcare: AI can be used to analyze large amounts of medical data and help doctors make more accurate diagnoses and treatment plans.
4. Safety and security: AI can be used to enhance security and public safety by identifying and preventing potential threats, such as fraud or terrorism.
5. Bias and ethics: AI systems can perpetuate biases and discrimination if they are not designed and trained in an ethical manner. Therefore, it is essential to ensure that AI systems are fair and inclusive.
6. Privacy: AI systems collect vast amounts of data about individuals, raising concerns about privacy and data protection.

Overall, AI has the potential to bring significant benefits to society, but it is crucial to manage its implementation and ensure that it is used in an ethical and responsible way.

Existing attempts to govern AI technologies reflect these difficulties. The first Responsible AI in the Military summit in the Netherlands in February 2023 was criticized as a [missed opportunity for US (and Chinese) leadership](#), where states agreed to develop AI "responsibly" but did not define what responsible use is. While European businesses [welcome the regulatory clarifications](#) accompanying the EU's AI Act, which is expected to come into force by the end of 2023, survey data suggests they also fear its implementation will [harm the competitiveness of Europe's AI industry](#). Previous attempts to regulate the use of autonomous weapons have been repeatedly stopped due to disagreements over [what "autonomous" means](#) in practice.

Clearly, exploring both the social implications and governance opportunities (domestic and international) afforded by AI is both timely and necessary. Over the next few weeks, the contributors to this commentary series will illustrate the interaction of both in diverse issue areas (e.g., economy, environment, culture, etc.) and governance settings (e.g., national, regional, global). My interview partners, in turn, provide insights into how these issues interact with goal development and implementation involving AI technologies in the private and non-profit sectors.

To maximize the breadth of our discussions, I adopt a broad [definition of AI](#) in this series, namely, "technology that automatically detects patterns in data, and makes predictions on the basis of them." I also acknowledge that technological progress may ultimately demand a re-evaluation of this definition. For instance, governance discussion of autonomous transportation systems call for "[humans in the loop](#)," essentially conceptualizing humans and technologies in a partnership as opposed to inventor-invention or governor-governed.

I and the contributors to this series look forward to such developments and to stimulating interesting discussions of this technology and its governance via this commentary series, both here and in the future. We also welcome your feedback and engagement, so please feel free to contact us via social media and email.

Laura Mahrenbach is an adjunct professor at the School of Social Sciences and Technology at the Technical University of Munich. Her research explores how global power shifts and technology interact, with a special focus on the implications of this interaction for the countries of the Global South as well as related governance dilemmas at the national and global levels. More information available at www.mahrenbach.com.

1. On the Normative Importance of the Distinction Between 'Governance of AI' and 'Governance by AI'

Eva Erman and Markus Furendal

In an era where increasingly complex and capable artificial intelligence (AI) systems are unveiled at a steady pace, the effects that AI technology may have on economic, social, and political issues become increasingly clear. While many tasks in blue-collar jobs have already been automated, for instance, knowledge workers have generally been considered to perform creative tasks that machines are unable to recreate. Yet, recent advances in 'generative AI' technologies that instantly create text or images based on short prompts, have left illustrators, writers and office workers reconsidering their job security.

Although new technologies such as AI are sometimes thought of as self-propelling forces, their social and normative implications are not predetermined but rather a result of political decisions and dynamics. In light of the anticipated and actual social impact of AI technology, calls for AI governance are thus more common than ever. Yet, even though the term AI governance is widely adopted, it is still largely undertheorized, and frequently used to describe a variety of distinctive phenomena and ideas.[i] Our aim in this short text, which draws on arguments we have presented at length elsewhere, is to introduce a helpful distinction which may reduce the risk of misunderstanding, and enable researchers and the public to better think through the role of politics in the age of AI.

Governance of AI and governance by AI

In the public and academic debates on the social impact of AI, the term AI governance is often used to refer to two phenomena that we suggest are in fact distinct: the phenomenon of 'governance of AI' and the phenomenon of 'governance by AI'. The former term refers to the kinds of emerging governance structures at various levels of policy-making that regulate and steer AI development and deployment. The most relevant example is perhaps the EU's so-called 'AI Act', which some expect will move from the draft stage into a binding regulation before the end of 2023. The latter, by contrast, describes the phenomenon of institutions implementing AI systems into their existing governance mechanisms. Many public authorities, for instance, already rely on AI systems to process data, automate decision-making, and detect suspected fraud. When public agencies do this, they govern by, or at least with the help of, AI.[ii]

Governance of AI includes hard law such as the coming EU AI Act, but also countless efforts best described as soft law approaches. These include recommendations, standards, ethical guidelines and declarations, codes of conduct and similar instruments developed by AI companies, NGOs, international organizations, or other actors in the AI space.[iii] Given that the soft law approach is agile and that its instruments can be adopted even when there is little international cooperation and agreement, soft law makes up a substantial share of global AI governance, and many expect it to remain the dominant approach.

Governance by AI could perhaps also be described as more or less soft, depending on its character and effects. On this view, we are in a sense softly governed by the recommendation algorithms or customer service chatbots that we encounter in our daily lives, whose guidance we are ultimately at liberty to turn down. But we are also governed in another, more consequential way by, for instance,

private insurance companies that calculate algorithmically derived risk profiles, and public authorities that employ automated decision-making about crucial issues like access to welfare benefits. In many cases, decision-making is supported by – rather than outsourced to – machines, such that there is still a human in the loop, who formally makes the decision recommended by AI technology. Some suggest that we could go further, however, and altogether hand over decision-making to machines. Optimists suggest that the AI-driven data analytics tool could collect citizens' views and thereby extend and equalize political influence. The developers of the AI chatbot 'Politician Sam', for instance, claim that it can analyze social media to accurately capture the political views of voters, and promise that it can thereby deliver 'true representation', 'active engagement', and 'better policy'.

Although these two notions risk being conflated by the widespread use of the monolithic term 'AI governance', we argue that it is important to keep them distinct, not least if we consider AI governance in relation to key normative ideals, such as democracy. The reason becomes apparent once we ask what it means for AI governance to be politically legitimate. Elsewhere, we have developed an account of the political legitimacy of AI governance, which attaches significance not only to the outcomes of, but also the procedures in, governance. Applying this account to actual AI governance suggests that both governance of AI and governance by AI can be politically legitimate under certain circumstances, but that these circumstances differ.

The political legitimacy of AI governance

We argue that the governance of AI is not necessarily politically legitimate when and because it produces 'good' outcomes, i.e. realizing the benefits and avoiding the risks of AI development. It also matters how we have come up with such a list of benefits and risks, and the goals of AI governance more broadly. Specifically, this process has to live up to some minimum threshold of democracy, where those who are affected by the decisions have an opportunity to participate in their making as equals.

To illustrate this point, consider the process of developing the EU's AI Act, on the one hand, and the AI-ethical work inside an AI-developing company, on the other. The EU's efforts in AI governance seeks to promote "trustworthy AI", an ideal which presupposes respect for human autonomy, prevention of harm, fairness and explicability. Similarly, the large company Microsoft is committed to promoting what it calls "responsible AI", which is assumed to include values like fairness, reliability, and privacy, as well as inclusiveness, transparency, and accountability. At the face of it, it is very difficult to tell these somewhat vague ideals apart, and it is an empirical question which of the two efforts will ultimately be most significant. The EU is hoping for a 'Brussels effect', where legislative action in Europe sets the standard for the rest of the world. On the other hand, Microsoft is such a dominant player in the AI sector that their internal guidelines might very well be more consequential for the future of AI development.

Our point, however, is that aside from the actual effects of AI governance, it also matters whether there is a 'chain of legitimacy' between those who make decisions, and those who are affected by them. From this perspective, the key difference is that the EU's rules are legitimate, since they can ultimately be traced back to EU citizens, while private companies like Microsoft exercise authority that lacks this kind of legitimacy. This conclusion follows from an argument that we have made elsewhere regarding the global governance of AI, which we will summarize here.

Currently, most attempts to steer AI development toward certain outcomes are initiated either by private actors such as Microsoft, or by public entities. The latter can be better understood by distinguishing between two ways in which citizens can give public institutions the right to rule. First, authorized entities have been granted power by citizens through a direct authorization. You authorize your nation-state's parliament, for instance, when you go to the ballot box and elect a representative. In the AI space, authorized entities set up the legal structure for the societal goals and overall aims of AI development and deployment as well as the basic form of the main institutions of the AI space, through coercive decision-making. These institutions have the right to rule because they have been established through a democratic procedure in which those affected by AI (in one form or the other) have had an opportunity to participate as equals in shaping the "control of the agenda" concerning AI.

Mandated entities, by contrast, have been delegated political power not from citizens directly, but from authorized entities. They make non-coercive administrative decisions, work out policies, and so forth. You are governed by a mandated entity when you, for instance, follow rules set out by executive bodies or interact with public administrative agents when applying for benefits or permits. A mandated entity can also further delegate authority to another mandated specialized entity. It may be appropriate to do so if, for instance, it enables higher-quality, decentralized and specialized governance.

Our conclusion about the difference between the EU's and Microsoft's AI governance follows from the assumption that democracy presupposes that instances of authorization and delegation constitute a legitimacy chain between those affected and the decision-making entities. The governance of AI is hence legitimate only when there is such a legitimacy chain. Even though the EU is often accused of having a 'democratic deficit', there is nevertheless a formal democratic connection between individual EU citizens and the institutions in which this law is taking shape. The AI Act was first proposed by the European Commission, whose legitimacy can be traced back to the citizens of EU through a chain of authorization and delegation. By contrast, the soft law approach spearheaded by non-authorized, non-mandated tech companies like Microsoft lack this kind of legitimacy. Regardless of how laudable their aims are and how efficient a soft law approach is, there would thus be a legitimacy deficit if these initiatives were the only kind of governance of AI.

Legitimate governance by AI?

The distinction between governance of and governance by AI is also significant because it helps us understand what is going on in cases like the AI-Politician Sam. AI systems are often described as having superhuman capacities to gather, analyze and summarize data. In general, decisions are handed over to AI systems precisely because we believe they are better than humans at identifying the right option. We suspect that some will think that it makes sense to hand over many of the decisions that are currently made in a democratic fashion to AI systems, if it would lead to better outcomes (whatever that is taken to mean). If that happens, then it appears that an AI system would be an authorized entity, wielding legitimate authority over its human subjects.

In our view, however, governance by AI systems in this stronger sense cannot be legitimate. This is because democracy is not merely a decision-making method to reach good decisions, but also an ideal of self-determination, according to which those who are supposed to comply with the rules have had the opportunity to authorize them by participating in their making as equals. There is no

principled reason why mandated entities such as public administrations cannot legitimately engage in governance by AI, such as AI-assisted decision-making in relation to a predetermined set of issues within an already established legal framework, like when to grant or deny applications for welfare support.[iv] And AI systems could perhaps even become a kind of mandated entities, if authorized entities delegate some power to them. Human decision-makers in a parliament could, for instance, rely on AI-based technology to make more informed and thus better decisions.[v]

Yet, since the aims and goals of political communities ought to be deliberated and decided upon collectively, by the people bound by rules and regulation, we find it difficult to defend the claim that an AI system could also be the ultimate source of political legitimacy, i.e., that it could be an authorized entity. On our view, even if such a hypothetical AI agent would provide better decisions, handing over authority to it would negatively impact political legitimacy as we have conceptualized it.[vi] Given the speed at which AI systems are currently developing, however, we believe this is a key issue for future research.

Conclusion

In conclusion, we believe that much could be won by researchers and the public paying closer attention to the ambiguous character of the concept of 'AI governance'. Moreover, both of the phenomena we have described here raise substantial normative and practical questions about the way in which politics and AI technology interact. We have begun to describe here – in much-simplified and broad terms – some of the considerations one should keep in mind when considering the political legitimacy of AI governance. Given the wide-ranging and deep effects that the advent of AI technology is likely to have on societies world-wide, it is crucial to continue to study and develop theories for when and how the governance of AI, as well as governance by AI, live up to the ideals that should characterize people's social and political interactions.

Eva Erman is Professor at the Department of Political Science, Stockholm University. She works in the field of political philosophy, with special interest in democratic theory, critical theory and philosophical methodology. Erman is the author of The Practical Turn in Political Theory (2018) and Human Rights and Democracy: Discourse Theory and Global Rights Institutions (2005), and has published numerous articles in scholarly journals, including British Journal of Political Science, The Journal of Politics, and Political Studies. Furthermore, since 2008, Erman is the founder and Editor-in-Chief of the journal Ethics & Global Politics (Routledge).

Markus Furendal is a postdoctoral researcher at the Department of Political Science, Stockholm University. He focuses on issues concerning the social and ethical impact of AI, distributive justice, and the future of work. His work has been published in journals such as Political Studies, Social Theory & Practice, Philosophy & Technology, and Journal of Applied Philosophy.

[i] See, for instance, Allan Dafoe, "AI Governance: A Research Agenda". Governance of AI Program, Future of Humanity Institute, University of Oxford, (2018), Anna Jobin, Marcello Ienca, and Effy Vayena, "The Global Landscape of AI Ethics Guidelines," Nature Machine Intelligence 1, no. 9 (September 2019): 389–99, , and cf. B. Guy Peters, "Governance As

Political Theory," in The Oxford Handbook of Governance, ed. David Levi-Faur (Oxford: Oxford University Press, 2012)

[ii] We do not claim that this distinction is entirely novel. For instance, researchers have used the term "governance by algorithms" to describe the impact of recommendation algorithms on people's construction of a social order. Natascha Just and Michael Latzer, "Governance by Algorithms: Reality Construction by Algorithmic Selection on the Internet," Media, Culture & Society 39, no. 2 (March 2017): 238–58. See also Kuziemski Maciej, and Misuraca Gianluca. 2020. "AI Governance in the Public Sector: Three Tales from the Frontiers of Automated Decision-Making in Democratic Settings." Telecommunications Policy 44 (6), and Christian Katzenbach and Lena Ulbricht, "Algorithmic Governance," Internet Policy Review 8, no. 4 (2019). Further, the term "algocracy" has been used to refer to something akin to what we here label governance by AI, by John Danaher, "The Threat of Algocracy: Reality, Resistance and Accommodation," Philosophy & Technology 29, no. 3 (2016): 245–68.

[iii] Thilo Hagendorff, "The Ethics of AI Ethics: An Evaluation of Guidelines," Minds and Machines 30, no. 1 (March 2020): 99–120, https://doi.org/10.1007/s11023-020-09517-8; Jobin, Ienca, and Vayena, "The Global Landscape of AI Ethics Guidelines."

[iv] There might, however, be practical reasons to be wary of this, relating to the explainability, accuracy and underlying fairness of the predictions and decisions generated by AI systems. For discussions, see Renée Jorgensen, "Algorithms and the Individual in Criminal Law." Canadian Journal of Philosophy 52, no. 1 (2022), 61-77; Kaun, Anne. 2022. "Suing the Algorithm: The Mundanization of Automated Decision-Making in Public Services through Litigation." Information, Communication & Society 25, no. 14: 2046–62; Kate Vredenburgh, "Fairness." In The Oxford Handbook of AI Governance, edited by Justin B. Bullock, Yu-Che Chen, Johannes Himmelreich, Valerie M. Hudson, Anton Korinek, Matthew M. Young, and Baobao Zhang. Oxford: Oxford University Press, (2022); and Liesbet van Zoonen, "Data Governance and Citizen Participation in the Digital Welfare State." Data & Policy 2 (2020): 10–17.

[v] The problems mentioned in the previous endnote are relevant also in this scenario.

[vi] One objection to our view could say that people can authorize an AI system to make future decisions in their place, just like members of parliament are authorized to make decisions in representative democracies. Yet, this once again implicitly assumes that authorized entities implement an agenda that is already set. Our point above is that the content and shape of the democratic agenda ought to be deliberated and decided upon collectively, by those bound by it. For an alternative line of argument reaching a similar conclusion as ours, see Ludvig Beckman, , Jonas Hultin Rosenberg, and Karim Jebari. "Artificial Intelligence and Democratic Legitimacy. The Problem of Publicity in Public Authority." AI & Society, (2022), who suggest that it is the opacity of machine learning undermines the publicity necessary for the AI to have legitimate authority.

2. Do more clubbing. How forming partnerships can help tech middle powers survive the escalating US-China tech war

Maximilian Mayer and Gedaliah Afterman

As the rivalry between the United States and China intensifies, with a growing emphasis on technology, middle powers are being increasingly caught between two superpowers. The strategy that these governments choose to adopt could have significant ramifications. Many countries are turning to various forms of hedging strategies. In line with a traditional middle power pattern, countries such as the Philippines, the UK, Singapore, and Vietnam have been forging closer defense ties with the US while maintaining and even strengthening economic cooperation with China.[i] Japan, the UK, and Australia have recently moved closer to a bandwagoning model, aligning their economic and security policies with the US. However, most leaders reject the idea of having to choose between the US and China. As Singapore's Prime Minister Lee Hsien Loong [stated](#), "It is not possible for us to choose one or the other, because we have very intense and extensive ties with both the US and China."

This analysis seeks to highlight some of the current trends in superpower competition and their impact on middle countries. We will ask how middle powers can manage the shrinking room for hedging strategies[ii] and maintain autonomous foreign and technology policies which are becoming ever more entangled due the growing geopolitical pressure to take sides.

Current state of play

The strategies that the US and China employ to compete are the most consequential examples of weaponized economic interdependence, where one party uses [interdependence to coerce](#) or deter the other rather than seek to cooperate for mutual gain. What initially was a bilateral issue between the US and China has through laws and presidential decrees regarding tech exports sanctions and further restrictive regulations become an increasingly dominant global phenomenon. Middle powers are struggling with this new era in which powerful governments use technological and economic choke points to exert pressure on rivals and friends alike.[iii] But the reordering process currently underway has wider implications, especially for the ability of international organizations such as the World Trade Organization to shape collective action. The contraction of globalization greatly impacts the economies of small and middle countries and especially their technological development, as the competitiveness of firms depends on market access, the flow of expertise as well as transnationally linked innovation processes.

The US is employing deft economic statecraft for dealing with China. Starting with the Trump administration, Washington has been working to put trade and technology sanctions in place against Beijing. The Biden Administration's tough policies regarding China are

seemingly the only issue that enjoys bipartisan support in a highly polarized US Congress. 5G networks and advanced computer chips are examples of a digital infrastructure and a core technology, respectively, which are becoming [the central battleground](#) on which China and the US are competing. Having shifted from viewing China first as a competitor, then as a challenger, and now as a threat, the US has chosen to pursue an aggressive containment strategy, that aims at "slowing down" the speed of Chinese innovation and freezing China's digital tech base at its current level thereby substantially hindering China's ability to develop in many areas.

While these policies aim at curbing further growth of China's economic and military might, their effectiveness depends on the cooperation of important US allies. When it comes to cutting edge semiconductors, high-tech companies from key US allies South Korea, Japan, Germany and the Netherlands control some of the most critical technological components.[iv] As their enthusiasm to follow US tech restrictions against China has been lackluster so far, the Biden administration has started to resort to strong measures such as the extraterritorial application of US law to supplement already vigorous diplomatic efforts in order to get allied countries in line.

China, for its part, is striving to develop a more autonomous domestic tech industry and exerting pressure on US allies to stave off further technological isolation. As it reenters the world stage, after three years of COVID-enforced isolation, China is also offering incentives such as the promise of tech cooperation and investments. The most recent examples are proposals for renewable energy investment in the Philippines and space cooperation with Djibouti. When Philippine President Ferdinand Marcos Jr. was in Beijing on January 5, 2023, nine Chinese energy companies [pledged to invest](#) $13.76 billion in the Marcos administration's push for renewable energy. A week later, two private [Chinese space companies](#) signed a memorandum of understanding with the government of Djibouti to "build a $1 billion commercial spaceport with seven launch pads and three rocket engine test facilities."

The rise of minilateralism

For middle powers, it is crucial to develop the skill of anticipating and responding effectively to pressures from larger nations. One example is the provisions in the recent US Inflation Reduction Act (IRA) that favour production in the US, which caught many of the country's allies by surprise. The domestic policies of the US have far-reaching effects, impacting innovation, science, and industrial policies globally. The globalization-based multilateral trading system is at risk of becoming determined by the foreign and security policy priorities of the US and China rivalry.

To that end, several middle countries have added a more flexible strategic approach to their toolbox that complements traditional bandwagoning, balancing and hedging: minilateral connectivity diversification. Minilateralism has been [described](#) as bringing "to the table the

smallest possible number of countries needed to have the largest possible impact on solving a particular problem". Empirically, it comprises small and medium-sized countries increasingly cooperating through clubs and other forms of small, flexible groupings, furthering the trend towards multipolarity. New trade and investment agreements, as well as collaborations in technology and research, play a vital role in maintaining integrated markets and protecting against fragmentation in the technology sphere.

Some of the origins of minilateralism lie within the global system. Governments are moving to establish bespoke intergovernmental cooperation due to the demise of multilateral organizations which suffer from a loss of efficacy – two interlinked trends that are comparable to the blossoming of minilateralism during the final phase of the Bretton Woods system.[v] Yet a second feedback loop is at play too. While the attractiveness of minilateralism partially results from the decline of US economic power and its ordering function,[vi] its effects at the same time render the superpower's ability to translate power resources into outcomes less effective. Neither the US nor China can completely dominate the agenda. Each superpower tries hard to convince its smaller allies to adopt policy approaches supportive of its own national interest.   Neither of them has been successful in this endeavor so far. Germany is not prepared to give up on the Chinese export markets so crucial to its economic strength. South Korea does not whole-heartedly embrace the economic decoupling strategy against China propagated by the Biden administration. And Israel finds itself trying to maneuver between growing pressure from the United States and its interest to maintain economic cooperation with China, especially on technology cooperation.

What Should Middle Powers Do?

In this situation, small and medium-sized countries should adhere to five guiding principles for shaping future foreign and technology policies.

#1 Form and use minilateral groups to realize your national interest.

The current global system is likely not destined to become polarized as it was during the Cold War. Today's international environment is more complex, more interconnected and more dynamic. The most critical insight stemming from this perspective is effective collective action does not require a hegemon or a strong institutional framework. The fact that the Transpacific Partnership (TPP) did not collapse but was implemented as a Comprehensive and Progressive Agreement for Trans-Pacific Partnership (CPTPP) after the US withdrew, is a primary example of the agency of middle powers, and collective leadership in action.

#2 Become a member of many clubs.

Diversification of the portfolio of minilateral partnerships is not only an effective risk insurance but also creates more foreign policy options. Having access to the strengths and

knowledge of other countries facing similar problems transcends the Cold War "non-aligned" mentality and is a key to success. Take, for instance, the case of Germany's energy policy. As Berlin aims to secure Green Hydrogen supplies, it should work to create a group of likeminded countries rather than seeking to conclude bilateral agreements. Germany's bitter experience with Nord Stream — bilaterally implemented with Russia — illustrates the costs of misplaced bilateralism. Indeed, as the recent example of an Israeli-Egyptian-European gas export partnership shows, when middle powers operate through new and, sometimes unorthodox, groupings, it can give rise to surprising breakthroughs.

#3 Form issue-based and interest-based clubs following the domestic-global gradient.

Creating new clubs that are helpful in achieving foreign policy goals and promoting overlapping domestic priorities can be highly beneficial. Frequently, domestic agendas are tied into global connections and processes. Yet, policymakers who are primarily focused on their own national agenda tend to underappreciate these connections. As a result, many opportunities to realize synergies regarding, for instance, energy transition, digitalization and innovation policies, are missed. The "Climate Club" recently established by the G7 exemplifies how states try to use minilateral formats to better coordinate their climate-related policies outside existing multilateral frameworks. Geography should not limit or determine the composition of such groupings. Transregional clusters are forming to address joint challenges as the growing cooperation between India, Japan and the Republic of Korea with countries in the Middle East in areas such as energy and food security and technology indicates.

#4 Operate minilateral groups pragmatically and informally.

Minilateralist practices tend to be more flexible and less formalized than multilateralism to put aside, as far as possible, strategic differences, hierarchies and formal procedures. This is evident in the I2U2 grouping where, despite different strategic outlooks regarding issues such as China, the war in Ukraine, and the Iran nuclear issue, the four partners (India, Israel, the UAE, and the US) have agreed to mutually beneficial economic and technological cooperation. The cases of BRICS (Brazil, Russia India, China, South Africa), Quad (USA, Japan, Australia, India), ASEAN Plus 3 and Shanghai Cooperation Organization (SCO) show that overlaps across the economic and security fields are possible. Similarly, involving Japan, South Korea and even China into the context of the Abraham Accords between Israel and its regional partners would make sense. In practice, minilateral statecraft is most effective when avoiding the trap of playing "exclusionary" [vii] games directed at the US, China or other actors.

#5 Form or join minilateral clubs to mitigate harmful central control and tech monopolies.

In the digital age, with its monopolies and structurally designed digital dependence regarding payment and data platforms, technological infrastructures create huge global

asymmetries. These affect specifically middle power autonomy in radical ways. For instance, in September 2020, Facebook threatened to block Australian users from sharing local and international news in a row with the Australian government over new media legislation. Therefore, it makes sense for middle powers to avoid over dependence on US/China social media, platform giants and infrastructure. In addition, as the example in the European Union has repeatedly demonstrated, tech middle powers often need partners to negotiate successfully with global big-tech firms such as Meta, Google or Apple.

Stabilizing the world for the "rest"

In the aftermath of Russia's invasion of Ukraine, middle powers are under pressure to respond to the primacy of geopolitics over economic globalization. For middle powers, the sound answer to a world increasingly impacted by weaponized interdependence may seem counterintuitive: create more connectivity, not less, and use it cooperatively for mutual gain. The key to a multipolar world and to avoiding getting caught up in superpower faceoffs is multidirectional diversification rather than decoupling and isolation. Middle powers can influence and shape events in a challenging context by forging new alliances and tech cooperation and resisting succumbing to the alluring siren's song of self-reliance and autarky.[viii]

Middle powers can collaborate to establish a more stable and interconnected environment, even as Beijing and Washington may be headed towards a collision course. By forming new technology and trade partnerships, middle powers can not only strive to reduce the risks of asymmetric dependencies, but also shift the global discourse away from a focus on systemic competition and the dominance of military power. The rest of the world should not wait for the two major powers to end their dangerous dance, as they must address pressing issues such as managing energy transitions, fostering innovations for sustainability, and addressing social inequality.

Maximilian Mayer is Junior-Professor of International Relations and Global Politics of Technology at University of Bonn. His research interests include the global politics of science and technology; China's foreign and energy policy; global energy and climate politics; theories of International Relations. He has authored China's Energy Thirst: Myth or Reality? (2007 together with Xuewu Gu), Changing orders: transdiciplinary analysis of global and local realities (2008, co-editer), two-volumes on The Global Politics of Science and Technology (2014, lead editor). Maximilian is coeditor of Art and Sovereignty in Global Politics (Palgrave, 2016), edited Rethinking the Silk-Road: Chinas Belt and Road Initiative and Emerging Eurasian Relations (Palgrave, 2018) and is currently leading the research group on „Infrastructures of China's Modernity and Their Global Constitutive Effects", funded by the Ministry of Culture and Science of the State of North Rhine-Westphalia.

Gedaliah Afterman is head of the Asia Policy Program at the Abba Eban Institute for Diplomacy and Foreign Relations at Reichman University (IDC Herzliya), Israel. He has written extensively on Asian regional security, Chinese foreign policy, and issues related to superpower competition and growing cooperation between middle powers in Asia and the Middle East following the Abraham Accords. He has a particular interest in China's evolving role in the Middle East. Gedaliah previously served as an International Strategic Policy Specialist at the Australian Department of Foreign Affairs (DFAT) focused on Asian regional security and strategic issues and as a diplomat at the Australian Embassy in Beijing, where he worked on issues related to China's foreign policy. He was previously Fellow and China Program Lead at the Jewish People Policy Institute (JPPI) in Jerusalem. Gedaliah holds a B.A. in Philosophy from the Hebrew University of Jerusalem and a PhD from The University of Melbourne.

Photo by [Aleksandar Pasaric](#)

Notes

[i] Enrico Fels, 2017, Shifting Power in Asia-Pacific: The Rise of China, Sino-US Competition and Regional Middle Power Allegiance. Switzerland: Springer International Publishing.

[ii] Smith, Nicholas Ross. "New Zealand's grand strategic options as the room for hedging continues to shrink." Comparative Strategy 41.3 (2022): 314-327.

[iii] Leonard, Mark. Connectivity wars: Why migration, finance and trade are the geo-economic battlegrounds of the future. European Council on Foreign Relations (ECFR), 2016; Drezner, Daniel W., Henry Farrell, and Abraham L. Newman, eds. The uses and abuses of weaponized interdependence. Brookings Institution Press, 2021.

[iv] Miller, Chris. Chip War: The Fight for the World's Most Critical Technology. Simon and Schuster, 2022.

[v] Orfeo Fioretos (2019) Minilateralism and informality in international monetary cooperation, Review of International Political Economy, 26:6, 1136-1159.

[vi] Brummer, Chris. Minilateralism: how trade alliances, soft law and financial engineering are redefining economic statecraft. Cambridge University Press, 2014.

[vii] See: Brummer, Chris. Minilateralism: how trade alliances, soft law and financial engineering are redefining economic statecraft, p. 20.

[viii] See: Helleiner, Eric. "The return of national self-sufficiency? Excavating autarkic thought in a De-Globalizing Era." International Studies Review 23.3 (2021): 933-957.

3. Contextualizing China's AI Governance

Jing Cheng

Artificial intelligence (AI), perceived as a revolutionizing force, has drawn mounting attention worldwide. China is noticeably striving for AI leadership, transitioning from "a norm-taker towards a norm-shaper, if not maker". China's approach to AI governance is often regarded as starkly different from the European and American approaches, with many regarding it as state-led and "the vanguard of digital authoritarianism" and highlighting the AI race between China and US. Here, I suggest a more nuanced approach to discussing China's AI governance policies and practices, contextualizing these within the Chinese sociocultural context, and identifying some challenges ahead. I argue that regardless of differences, a contextualized, inclusive framework – one that considers the domestic dynamics and brings stakeholders and countries together – is beneficial for building a common digital future.

AI governance policies and practices

China's aspirations to become a great power in science and technology are closely tied to its historical experience about technological backwardness, often quoted in the discourse as "the backward will be beaten." China consequently aspires to develop science and technology to transform the country into a sci-tech powerhouse. In the digital realm, China has invested substantially in digitalization and modernization. AI as an emerging technology offers China an opportunity to potentially lead the world in AI theories, technologies and applications.

Given that AI is ubiquitous in its applications, ranging from recommendation and navigation to general-purpose chatbots, there are many regulatory needs to monitor and regulate AI. In China, a national AI plan was issued in 2017, and it set goals for achieving an AI governance system by 2030. Since then, various documents have been released, including China's Civil Code (2020), Personal Information Protection Law (2021) and Ethical Norms for New Generation AI (2021). China has recently stepped up to propose draft measures to regulate the use of generative AI. These regulations – often issued by different AI-related bodies – reflect China's strong impulse to regulate the privacy, data security and ethics aspects of AI.

China has also published guidelines and rules for AI standard setting. These include Guidelines for the Construction of a National New Generation Artificial Intelligence Standards System (2020) and Artificial Intelligence Standardization White Paper (2021). There are different levels of standardization at the national, industrial and enterprise levels. This is in part a reflection of the Chinese AI landscape, in which multiple stakeholders at different levels get involved in the agenda-setting and decision-making process to regulate AI.

Another notable aspect is governance by AI, an issue of importance to the Chinese government. Good governance (shanzhi) is increasingly associated with "smart government" (zhihui zhengfu), highlighting the use of technology in providing better public services and supporting better planning and decision making. Studies show that responsive e-government and providing convenient and efficient services are positively received by the public. In China, AI applications and facilities are promoted in public service systems, such as traffic and court systems, as a means of optimizing public services for citizens. One example involves a series of smart city development and governance projects carried out under partnerships between the government and tech giants such as Baidu, Alibaba and Tencent. The Hangzhou government and Alibaba have jointly built the City Brain system of urban traffic management. It helps city administrators analyze live streams of traffic and improve

the incident identification accuracy rate, and it has been rolled out in at least 23 cities inside China and internationally, including in Kuala Lumpur in Malaysia.

The process of AI governance in China is not only top-down and based on a monolithic government masterplan, as is often assumed, but also involves interactions between the state and digital companies as well as other stakeholders, such as universities, research institutions and non-governmental organizations. Private digital companies make up national AI teams that promote AI innovation and shape the societal ecosystem. Provincial and municipal governments also try to connect multiple stakeholders for AI governance promotion and implementation. Peking University's Institute for Artificial Intelligence and the Optics Valley in Wuhan, for instance, have pledged to jointly build a smart social governance trial base, working with eight companies to integrate AI industries, education, research and applications.

It should also be noted that some AI governance principles being promoted in China stem from academia and industry, not necessarily from the government. An example is the Beijing AI Principles, which were released in May 2019 under the joint collaboration of the Beijing Academy of Artificial Intelligence, leading Chinese universities such as Tsinghua University, Peking University, the Chinese Academy of Sciences, and the Artificial Intelligence Industry Alliance (AIIA). In this evolving digital landscape, China's AI governance is and will continue to be shaped by a variety of stakeholders, including central and local governments, digital companies and academia, and their interactions.

Sociocultural context matters

Although technology itself is neutral and objective, at least theoretically, effective governance of AI and effective application of AI for governance needs to take local and national contexts into consideration, such as social norms and cultural traits. This is especially so when it comes to AI ethics.

In China, the often vague, abstract slogans in AI principles reflect Chinese philosophical and cultural practices applied to AI ethics. For example, the "Beijing AI Principles" call for healthy development of AI, highlighting the importance of harmony and cooperation "so as to avoid malicious AI race, to share AI governance experience, and to jointly cope with the impact of AI with the philosophy of 'Optimizing Symbiosis'". Optimizing symbiosis stems from the Confucian philosophy of harmony, underlining the harmonious existence among people and the symbiotic relationship between humanity and the environment, in this case the machine. Similarly, in Tencent's AI Principles, the concepts of "Tech for Good" and "digital well-being" (shuzi fuzhi) also highlight such a human-machine symbiosis, exploring "the balance between AI, individuals and society." In a global landscape of human-centred AI ethics, such philosophical understandings should be noted as an important context in which AI initiatives and principles are embedded locally and nationally.

Examples of how Chinese culture is reflected in China's approach to AI governance can be found in many AI-related documents, issued by industry or government organizations. One example is the draft Joint Pledge on Artificial Intelligence Industry Self-Discipline, released by AIIA. This document pledges to implement self-discipline and industry supervision mechanisms for AI ethics. In the AI Industry Responsibility Declaration, leading digital companies such as Baidu, Huawei and Ant Group jointly commit to pay great attention to the issue of social responsibility and to implement "self-discipline and self-governance" in AI. The Joint Pledge on Internet Information Service Algorithmic Application Self-Discipline, which is widely supported by 105 entities including industrial alliances, top digital companies and media, also attaches great importance to responsibility and self-discipline.

"Self-discipline" and "self-governance" have been highlighted as means of integrating ethical principles into all aspects of AI practices. The call for self-discipline in AI governance, also stated in the subsequent national [Ethical Norms for the New Generation Artificial Intelligence](#), is associated with the Chinese cultural tradition of self-cultivation and humanism. It draws on the Chinese philosophical theme of the prominent role played by the individual in social development and the inner transformation of oneself for better morality and governance in an ideal social system. In these Chinese approaches to AI governance, the concept of self-discipline and self-governance assign significant responsibilities for governing AI to governmental bodies, individuals and especially corporate actors.

Challenges ahead

As is evident, multiple stakeholders across China are shaping the emerging AI governance regime in China. China's AI governance nonetheless faces several challenges at the national level. One major challenge is the fragmentation of governance, with different layers of regulation and different bodies for AI governance. China has five levels of government administration. Different governmental bodies and digital actors, including tech giants and startups, have their own preferences and interests in promoting AI regulations, leading to internal struggles for resources, publicity, and influence. The landscape of China's domestic governance of AI is liable to be [as fragmented and decentralised as that of China's BRI projects](#).

Another challenge is the centrality of abstract philosophical concepts such as harmony and self-discipline as applied to AI regulation in China. The vagueness complicates AI ethics implementation and makes supervisory oversight challenging, whether self-governed at the corporate or individual levels or directed by the government. When pledges and guidelines are made one after another, abstract notions of self-discipline and self-governance would turn out to be ineffective for the regulation of AI. The ambiguity creates a situation where follow-up measures and monitoring are needed to ensure proper implementation. Otherwise the documents are more likely to produce empty talk than real effects.

An additional challenge arises from the interdisciplinary nature of AI itself. Although it is often stated clearly that an interdisciplinary perspective and joint efforts are essential, AI governance efforts in China often face difficulties when trying to implement interdisciplinarity and diversity in practice. Science and technology are highly valued in China and play a crucial consultation role in AI governance initiatives. Yet the perspectives from humanities and social sciences are also urgently needed in the formation of an effective and ethical AI governance framework for China. Relevant expertise is still relatively scarce in the AI governance discussions. Bringing the state, industry and all relevant branches of academia together to foster open and constructive discussion would help further improve the effectiveness of AI governance regime in China.

The conditions for AI governance, globally, are still evolving and far from clear. One area where China could make a positive contribution is in addressing existing inequalities at the global level, which have been further amplified in the age of AI. The Chinese government states that one goal of the digital transformation in China is to help [tackle the digital divide issue](#) worldwide, between the haves and have-nots. China could promote and support addressing under-representation in global discussions by focusing on groups that are most vulnerable, such as small and medium enterprises and technologically disadvantaged groups such as women and the elderly, at home.

Another area where China could contribute more is global cooperation. Despite global awareness of the emergent governance issues, a major global challenge is that geopolitical competition is growing over values attached to AI, digitalization, and the use and governance of advanced technology.

While technology is supposed to be neutral in an ideal scenario, humans nonetheless determine the functioning of the machine. Under the current international scenario, technology is increasingly treated as value-laden, and made into a geopolitical battlefield of value competition between technologically advantaged countries. Such competition further widens division and inflicts more problems. Some countries and groups – the technologically disadvantaged – are confronted with the dilemma of taking sides. This situation hinders the development of global governance of AI and could result in a dangerous global situation in which the lack of consensus and coordination in regulating AI leads more pressing AI ethical considerations that could affect sustained human life to be put aside.

More global cooperation on AI governance is needed, more inclusive frameworks are needed, as are joined efforts involving different segments of society and between nations. To do so, it is not helpful to simply bring "like-minded nations" together; it is rather more constructive for nations and groups to engage and interact, seeking common ground for AI governance while reserving differences in values.

Dr. Jing Cheng is a lecturer in the School of Foreign Studies and Deputy Director of Belt and Road Communication Research Centre at Xidian University. Her research interests lie in the field of national identity, Internet politics and international communication, and she is currently focusing on artificial intelligence and digital governance. She has published articles in journals such as Journal of Contemporary China and Journal of Asian and African Studies and also written op-ed articles for China Daily and The Conversation, among others.

4. Re-imagining Africa's sovereignty in a digitally interdependent world

Arthur Gwagwa and Beverley Townsend

As part of their AI strategies, the United States and China, with support from big technology companies, are pursuing goals to attain strategic autonomy and global dominance in cyberspace.[i] In an attempt to decenter U.S. and Chinese dominance, the European Union (EU) is pursuing the twin goals of attaining digital sovereignty and influencing the global AI regulatory landscape.[ii]

Scholars in the Global South and civil society have responded to this geopolitical positioning and the potential power asymmetries brought about through digital colonialism by re-focusing AI scholarship and advocacy on these emerging issues.[iii] But instead of joining hands with civil society and academics, certain African state actors are asserting their political interests by fervently appropriating calls for digital sovereignty and data localization. In this commentary, we discuss the dangers of domination associated with African pursuit of digital sovereignty and what role governance can play in meeting these challenges.

Domination under the guise of digital sovereignty

Some would argue digital sovereignty rhetoric is justified because the digital age will shape everything from commerce to international relations for the next few decades. In this context, the future of AI development and adoption and the prosperity of states will be (partly) decided by algorithmic control, computational capacity, a skilled workforce, and data ownership. Yet individual pursuit of self-determination can only be realised if the relations in which people stand to one another are non-dominating and are regulated both by the institutions in which they all participate and by ongoing negotiations among them.[iv]

It is thus important to determine how states and technology platforms shape the global AI governance framework, knowledge base, and the concepts that underpin them. Large online platforms are active in multiple African countries like Kenya, South Africa, and Nigeria, where they have a direct and significant impact on internal markets. This market dominance places them in a very powerful position by creating power asymmetries which impose barriers to entry for local corporations. Major states like China and the U.S. are also flexing their economic muscle. For instance, in addition to using a range of public diplomacy instruments to shape China's image in Africa, China is promoting an economic development narrative in Africa as a core component of its foreign policy there. Components include increased media engagement in Africa and projects developed via the Belt and Road Initiative. These projects, funded by Chinese banks and underwritten by the Chinese government, raise the possibility of unsustainable loans informing a so-called 'debt trap diplomacy.' Examples include a gas pipeline in Nigeria, the construction of a high-speed railway between Mombasa and Nairobi, and numerous projects in Uganda, Egypt, and Ethiopia. The difficulty is that China's strategy is not only to build but, more importantly, to control technological, information, and resource infrastructures.[v]

Simultaneously, major states, with the more or less ambiguous support of digital giants, are also shaping governance principles as part of their strategies for domination and/or autonomy in cyberspace.[vi] China's current hold on data infrastructures in Africa, means that China is a gatekeeper of consequential decisions in Africa's social and political realms. Also, in the political realm, U.S companies in Silicon valley wield tremendous power from a distance to determine public life in other countries. In the globalized economic system dominated by large online platforms, major decisions related to resource allocation are taken far from the individual producers and

consumers, and the web of causes and justifications is often complex. Further, the digital age is also reviving [geopolitical rivalry, in this case, between the U.S. and China](). Such rivalries have historically placed Africa at a disadvantage, with global superpowers engaging Africa as ["a theatre of operations rather than the focus itself]()."

Meanwhile, as previously mentioned, Africa national strategies include certain digital sovereignty measures. For example, several African countries have taken measures [to localize their data](). Following the launch of the country's first Chinese-built data centre on 22 June 2022, Senegal's President Macky Sall instructed his government to migrate all government data from foreign servers to a new national data centre. Similarly, certain African countries are pushing for regulations that impede the free flow of digital sequence information, which has [created ongoing tensions with the European Seed Association](). Such moves have generated multilateral discussions of so-called '[high-value data]()' and how it should, or should not, be shared. Some have argued that using African data without compensation ought to be restricted in an attempt to avoid the emergence of new digital power asymmetries, reminiscent of the colonial past.

Yet, while these are legitimate claims, the difficulty is that African appropriation of decolonization and digital sovereignty narratives is an assertion of domestic political interests as it is a reaction to global developments. For instance, certain ruling governments are re-purposing biometric data for [manipulative electoral campaigns and agreeing to massive data collection and management arrangements outside the national democratic accountability structures](). Political manipulation of "digital sovereignty" also creates domestic capacities for domination by national governments with the questionable support of foreign data brokers, for instance, by asserting a false sense of ownership and/or national security. This was evidenced recently in Senegal's Diamniado data center which the ruling government used [as a basis for claiming sovereignty]() when in reality it strengthened China's economic interests in the country. Particularly susceptible are countries with long histories of human rights abuses and poor records regarding the rule of law, like Zimbabwe and Angola. The combination of strong domestic opposition, ongoing insurgencies, and other security challenges, including popular protests, in these countries makes [China's surveillance technologies]() increasingly attractive to these governments. This has led to concerns that the pursuit of digital sovereignty can create opportunities for data privacy infringements – particularly, in countries with inadequate data protection laws, oversight, and enforcement.5 Others have gone farther, worrying that [political use of digital technologies worsens, rather than improves, existing governance challenges]().

The role of governance

Data concentrated in the hands of big technology companies necessitates 'taking back of control' and 'redressing power asymmetries' as crucial elements of African states' approach to AI and data governance. Encouragingly, national strategies for digital technology adoption and setting out paths for digital transformation have been developed and implemented in many African countries, including [Ghana](), [Kenya](), [Rwanda](), [Uganda](), and [South Africa](). At a continental level, the [AU's Digital Transformation Strategy]() builds on existing frameworks developed in Africa with the objectives of harnessing digital technologies and innovation to transform African societies and economies, generate economic growth, bridge the digital divide, and support socio-economic development. Similarly, the [Declaration on Internet Governance and Development of Africa's Digital Economy]() seeks to establish a transparent and inclusive framework of internet governance. This framework proposes principles, including freedom of expression, respect and protection of private life, universal access and technical interoperability, and the ethics and respect of cultural and linguistic diversity in cyberspace, as the basis for internet governance.

However, political fragmentation, a lack of strategic technology and data infrastructure autonomy, and a weak market position all continue to hinder Africans' position to influence global regulatory norms. As Christopher Gore writes about climate change negotiations, "The large number of independent African states and their diverse social, political, economic, and ecological conditions put the subcontinent at a collective disadvantage in bargaining, as it is nearly impossible for governments to present a unified position in negotiations." This is similarly evidenced in digital governance. African states have been slow to exercise collective political will and to move ahead with digital and data governance measures. This is reflected in the region's collective lack of political will to ratify treaties. Although the 14th ratification by the Democratic Republic of the Congo brings the number of ratifications closer to the required 15, it has been a long and slow process. The fragmented approach is also seen in the lack of a common approach to cybercrime. While certain African countries do not yet have cybercrime laws, other countries have ratified the Council of Europe cybercrime and data protection conventions. This fragmentation will affect Africa's capacity to influence global digital rules and will facilitate a weak market position in the global digitally interdependent economies, making it prone to domination by the powers that shape such rules. At the multilateral level, it is thus perhaps unsurprising that conversations on AI governance continue to reflect the perspectives of a few industrialized economies while the unique challenges of the Global South – from access to digital infrastructure and institutional capacities, to societal norms and cultural practices – remain outside governance discussions. For instance, a 2019 survey shows that, despite claims of universality, most AI ethics principles and their guidelines are developed by stakeholders based in economically developed, mostly Western countries like the U.S and those from within the EU.

What next?

There is an urgent need that perspectives and insights around digital and AI deployment in the Global South inform global governance narratives, or we stand to misalign AI expectations and requirements with the needs of a great many people of the world. We suggest the following:

- Engage in digital constitutionalism: Adapting core constitutional values to the requirements of the digital society in Africa can provide a scaffold upon which good AI governance and practice can be constructed. It reimagines and reframes rights and powers in the digital society and plays a pivotal role in Africa, where constitutional values are increasingly under threat by ubiquitous forms of surveillance, power, and control.

- Strengthen existing collaborations: African states, acting collectively though the African Union, should use their collective agency to diversify, but also strengthen, their respective existing collaborations with the EU in the area of cyber governance and with the U.S. on the inclusion of Africa's languages in the digital. Doing so can help build Africa's AI infrastructure, skills, and data management capacities. It can help Africans push for a governance framework that, like the EU data protection initiatives and AI regulatory proposals, are ethically desirable and normatively justified. And it can help Africans – acting through the African Union and with the assistance of the EU – to decenter the significant and influential position and presence China and Russia are adopting in the digital space in Africa.

- Create appropriate and contextually-relevant policies: Approaches to AI regulatory frameworks differ, sometimes significantly, between countries. (Take the different proposed approaches between the EU and the UK, for example). While core ethical values guiding policy frameworks on responsible AI can be shared between and across applications and locations, they must be implemented in accordance with particular contexts. As a result, some aspects of the principles may not apply – or will apply differently – in Africa. In this context, the African Union should continue to

pursue multilateral and intra-African approaches to AI governance which address local specificities, protect human rights, and advance digital reformation measures on the African continent.

- Promote inclusive governance: The United Nations and others can also assist through democratic channels that go beyond the current state-centric model proposed by China and Russia, by involving commercial enterprises and civil society to reframe and implement 'digital sovereignty' in Africa. Doing so is crucial for establishing useful governance principles and articulating how these values can be applied. It will also ensure that AI standards applied to Africa meet the democratic requirements of accountability. The role of accountability is vital and stands to verify compliance with substantive normative principles of good data and technology use including, amongst others, values of privacy, transparency, explainability, human oversight, and fairness.

- In conclusion, we propose a relational concept of the self-determination and non-dominance for African peoples. This, we suggest, aligns with African aspirations and claims in the era of global governance underpinned by digital interdependencies. It is also one more likely to accrue benefits for Africans in a time of increased digital and geopolitical struggle.

Arthur Gwagwa is based at the Utrecht University Ethics Institute and his research interests are on the geopolitics and intercultural ethics of AI.

Bev Townsend, PhD, is a South African researcher who works in the law and ethics of data and digital technologies, and in AI and resilient autonomous systems at the University of York, England.

Notes

[i] Theodore Christakis, 'European Digital Sovereignty': Successfully Navigating Between the "Brussels Effect" and Europe's Quest for Strategic Autonomy, Multidisciplinary Institute on Artificial Intelligence/Grenoble Alpes Data Institute, e-book, December 2020 Available at https://ssrn.com/abstract=3748098; https://airegulation.com

[ii] Anu Bradford, 'The Brussels Effect, How the European Union Influences the World', (2020)

[iii] Sabelo Ndlovu-Gatsheni, "Discourses of Decolonization/Decoloniality". Papers on Language and Literature, Vol.55, No.3, 2019, pp. 201-300. Available at https://www.blacfoundation.org/pdf/Ndlovu-Gatsheni-16.pdf; Abeba Birhane, 'Algorithmic injustice: a relational ethics approach', Patterns, Volume 2, Issue 2, 2021.

[iv] Iris Marion Young, 'Self-Determination as Principle of Justice'. Philosophical Forum_ (1979). 11 (1):30.

[v] Elenore Pauwels, 'Artificial Intelligence and Data Capture Technologies in Violence and

Conflict Prevention Opportunities and Challenges for the International Community', (2020) Global Center on Cooperative Security.

[vi] Christakis, 'European Digital Sovereignty'.

5. Can Green FinTech Build Climate Justice?

Laura Quinteros and Nick Bernards

Meaningfully addressing the challenges posed by climate breakdown requires massive investments both in reducing emissions and in adapting to a changing climate. So far, despite considerable global efforts, the actual mobilization of climate finance has fallen far short of expectations, particularly in poor countries. These failures raise important questions as to the allocation of finance, as well as subsequent distributional and procedural questions with respect to climate justice -- who should pay for climate mitigation and adaptation, and who should decide how mitigation and adaptation take place?

One increasingly common response to these failures has been to turn to new financial technologies with environmental functions -- or, 'green FinTech'. The label 'FinTech' bundles together an array of different mobile and digital technologies applied to the delivery of financial services. Common examples range from mobile payment systems, crowdfunding and peer lending platforms, and alternative forms of credit data through to cryptocurrencies, automated investment advisory services ('robo-advisers'), and online stock trading platforms.

'Green FinTech', by extension, is a loose term for the subset of these applications with expressly environmental aims. The Green Digital Finance Alliance (GDFA, 2022)-- a Geneva-based think tank launched by the UN Environmental Programme and Ant Group, gives the following definition: 'Green fintech solutions are defined as technology-enabled innovations applied to any kind of financial processes and products all while intentionally supporting Sustainable Development Goals or reducing sustainability risks'. Notable examples include pay-as-you-go (PAYGO) electricity systems combining mobile money applications with off-grid solar photovoltaic (SPV) power generation, various efforts to apply blockchain or crypto-token systems to carbon credits and other emissions offsets, crowdfunding for clean energy projects, the deployment of satellite data and artificial intelligence to screen and verify carbon offsets, and automated investment screening. In recent years, green FinTech has gained growing attention from both financial regulators and environmental agencies as a potential means of responding to shortfalls of climate finance.

Our aim in this commentary is to review the landscape of green FinTech and offer some critical discussion of its limits and possibilities. We see a core tension with green FinTech projects: The basic promise of green FinTech applications is that they will enable the screening and financing of 'green' projects 'at a distance'. They offer bundles of metrics, data, analytical tools, and payment infrastructures aimed at empowering investors to screen and verify 'green' projects quickly, cheaply, and remotely. Yet, these projects succeed or fail on their own terms (to say nothing of wider questions of climate justice) depending on how they are embedded with localized patterns of ownership, labour relations, and livelihoods -- precisely the complexities that, say, an AI programme scraping through satellite data is designed to remove. Moreover, they black-box vital and contested questions about how to reduce emissions or adapt to intensifying climate risks, and ultimately delegate decisions on those questions to software developers, investors, and automated programmes.

'Green FinTech': Merging climate finance and FinTech governance

Green FinTech bridges the landscapes of climate finance on one hand and the emerging governance of FinTech on the other. Both initiatives have come with some well-documented pathologies in practice.

Sarah Bracking and Benjamin Leffel point to the emergence of a regulatory architecture governing global climate finance which is increasingly polycentric, but also increasingly beholden to neoliberal logics privileging the interests of market actors. So far, the mobilization of climate finance through these arrangements has fallen well short of promises. As Table 1 shows, based on OECD data, the Paris Agreement pledge of USD 100 billion in climate aid annually has never come close to being met. Private finance was intended to provide a third of that 100 billion, but has only once reached half of that target.

Table 1: Climate finance provided or mobilized by donor countries, 2013-2019

	2013	2014	2015	2016	2017	2018	2019
Public bilateral (in USD billion)	22.5	23.1	25.9	28.0	27.0	32.0	28.8
Public multilateral (in USD billion)	15.5	20.4	16.2	18.9	27.5	29.6	34.1
Climate related export credits	1.6	1.6	2.5	1.5	2.1	2.1	2.6
Private finance (in USD billion)	12.8	16.7	N/A	10.1	14.5	14.6	14.9
Total	52.2	61.8	N/A	58.6	71.2	78.3	79.6
Private finance as percent of total	**24.5**	**27.0**	**N/A**	**17.1**	**20.4**	**18.6**	**18.7**

Source: adapted from OECD (2021).

Notes: no private sector data for 2015, as OECD implemented new measurement criteria, private finance figures from 2013-14 are not directly comparable to 2016-19

FinTech, meanwhile, has emerged as a key focus of financial regulators in recent years, particularly with respect to the promotion of 'financial inclusion' and poverty reduction. The World Bank and G20, together with a number of central banks and financial regulators in both Global North and South, have also increasingly promoted and coordinated targeted regulatory frameworks for FinTech applications aimed at promoting 'access' to finance for the 'unbanked'. A loose network of central bankers in particular have promoted 'regulatory sandboxes' -- time-limited, product specific licenses for particular companies to conduct 'experiments' with 'innovative' practices and technologies.

There are important parallels to the promotion of private climate finance visible here. The turn to promoting FinTech betrays a similar emphasis on market-based solutions to social problems, and on mobilizing private investment. In practice, this has meant that many of the same problems have appeared with FinTech applications as with private climate finance. The actual rollout of FinTech applications has been uneven, with heavy investment driven by readily available venture capital concentrated on a few key markets, notably Kenya and India, and on more profitable services, notably high-interest lending predominantly to urban-dwelling, 'less poor' borrowers.

The promotion of green FinTech brings many of the same regulatory tools to bear on the problem of climate finance. For instance, the Financial Conduct Authority in the UK has run two iterations of the 'Green Fintech Challenge' in 2018 and 2021 -- rolling out a 'regulatory sandbox' exercise specifically targeted to FinTech start-ups 'that will aid the transition to a net-zero economy'. The focus here, as with FinTech more broadly, is on creating an 'enabling environment' for FinTech experiments, in hopes of attracting private capital.

In sum, merging climate finance and FinTech regimes holds out the promise of breaking through some of the barriers to greater mobilization of climate finance. As we show in the next section using the example of PAYGO solar systems, this promise comes laden with significant tensions. The very features of green FinTech projects that make them potentially appealing to investors make them blind to important local dynamics which will determine their success or failure on their own terms, and threaten to undercut their viability as vehicles for climate justice.

Green FinTech in practice: Antinomies of PAYGO SPV electricity

In the context of the Global South, rural areas depict one of the main challenges for both public and private policies focused on universal energy access. Many rural households are scattered, have low and unpredictable incomes, and hence low energy demand. Central grid supply is thus often unprofitable for private suppliers and expensive for public ones facing fiscal constraints. Solar decentralized solutions are a key potential alternative for rural energy access, but the high upfront technology costs associated with both mini grid and stand-alone solar solutions remain a major challenge.

Against this backdrop, cutting-edge financial products relying on digital innovations are emerging and being deployed across different jurisdictions. For instance, PAYGO models coupled with mobile money for small-scale solar solutions are widely adopted energy access solutions in Southern countries. A digitally enabled PAYGO model allows users to pay for electricity in weekly, monthly instalments or when financially liquid using mobile payment platforms and enabled by machine-to-machine (M2M) technology incorporated in the solar solutions.

Many enthusiasts of digitally-enabled PAYGOs have been documenting the model's benefits to users. These include success in delivering affordable solar power and fair repayment performance according to a number of evaluations of projects in different parts of in Sub-Saharan Africa. Yet other authors have noted that the overall picture is mixed. In many cases, PAYGO solar systems appear to be profit-led and guided by market logics rather than guided by companies' supposed social vision. Measures of 'success' based on narrow measures of repayment rates and energy use risk missing out on key dynamics of power and exploitation. Lucy Baker describes the process as converting rural energy use into a set of financial assets grounded in new forms of consumer indebtedness.

Cross and Neumark document one such example: East Africa's digitally-enabled, off-grid solar power diffusion, an adverse 'infrastructure of inclusion' in which final users are governed by new circles of data, capital and debt. Data generates inputs for modelling and optimizing PAYGO business alternatives that enable new connections. But it also sets the grounds for disconnecting those defaulting on agreed payments. This is because the digital infrastructure can remotely lock out or shut down systems. The possibility of remote disconnection is significant for users and businesses alike given high rates of default on one hand and the notable material and social costs associated with repossessing SPV systems on the other. The costs of disconnection could also be immense for vulnerable populations. This is particularly true in COVID-19 and post-COVID 19 scenarios, whereby the loss of radio, TV or mobile phone to stay informed, or the loss of light in tandem with falling ill could be excruciating.

Moreover, the dynamics of indebtedness and distancing implicit in PAYGO solar systems may also create new ecological costs. Disconnection without repossession risks exacerbating the existing hazards from solar e-waste. Previous studies have illustrated how toxic materials contained in PV films and batteries threaten local ecosystems. And in fact, there are limited incentives for operators to reclaim or recycle disconnected SPV kits. Additionally, the intensification of indebtedness in agrarian settings has often been associated with intensified exploitation and depletion of water and soil resources, for instance, in Cambodia and India.

Conclusion

In short, digital solutions to the need for clean energy risk creating or exacerbating localized social and ecological risks. Moreover, they create these problems precisely because they are designed around the priorities of investors, with limited input from targeted communities and indifferent to localized dynamics of power and exploitation. These concerns ultimately challenge the operationalization of widely adopted frames in the climate finance discourse, including 'transformative change' and 'paradigm shift' according to which climate finance delivers regime-altering, new and transformative socio-ecological interactions in addition to inflows of capital.

Laura Quinteros is a Bolivian energy scientist and a PhD Candidate in Global Sustainable Development at Warwick University. She is investigating governance structures, rationalities and power relations that emerge in solar projects funded via crowdfunding platforms in the Global South.

Nick Bernards is Associate Professor of Global Sustainable Development at the University of Warwick. He is author of A Critical History of Poverty Finance (Pluto, 2022) and The Global Governance of Precarity (Routledge, 2018).

6. Who is the AI Owner? Hacking the Illusive AI Inventorship Challenge

Giuseppina (Pina) D'Agostino

Should a machine be treated like a human inventor and therefore be allowed to be an inventor on a patent? The paper posits that it is not so much who (human or machine or both) invents or even creates in the case of copyright, but who owns the intellectual property and who has the power to commercialize those assets that matters. Tackling the AI ownership challenge will more effectively ensure the public benefits from the intellectual property rights granted.

The Ownership Question

Dr. Stephen Thaler filed numerous patents worldwide for his AI inventing machine, called DABUS and invented precisely to invent. His filings spurned a significant question in the global intellectual property community: Does an AI inventor meet the various requirements for obtaining a patent? While important, the current debate among the global intellectual property community does not capture the entire challenge. Indeed, while salutary to answer the question of granting patents to an AI inventor, doing so will not solve the more pressing AI ownership challenge. Without considering the entire AI patent system and the humans who enable the commercialization of the AI patent, little benefit is derived for the public. And so, while there are many contrasting opinions on who should be the inventor and/or owner, there is little work on why ownership is a critical consideration.

The importance of the ownership question has applicability to the wider and quickly evolving AI industry worldwide. Indeed, generative AI, a new term to delineate the outputs of AI, is also subject to ownership issues and must be properly understood. While this paper focuses on AI patent ownership, it hopes to prove valuable to assess wider generative AI developments.

The Patent Bargain

The patent bargain is at the heart of the patent system, and any discussion that tackles inventorship and ownership in patents must depart here. The Supreme Court of Canada explained this bargain in Free World Trust v. Électro Santé Inc.: "In return for disclosure of the invention to the public, the inventor acquires for a limited time the exclusive right to exploit it." The Court elaborated on the concept of the patent bargain in Apotex Inc. v. Wellcome Foundation Ltd., describing a patent as "a method by which inventive solutions to practical problems are coaxed into the public domain by the promise of a limited monopoly for a limited time. Disclosure is what the public gets for allowing owners exclusive rights.

Central to the patent bargain is for society to benefit from the inventor's knowledge when the government grants a patent for an invention that is new, useful, and not obvious.[i] Through the work of patent examiners worldwide and the courts later testing their work – with much input from scholars and policymakers along the way – doctrines and principles are shaped over time to breathe life into this patent bargain. In one of his many foundational pieces, and long before it was trendy to write about intellectual property and artificial intelligence, David Vaver argued, "No two persons are likely to agree on [what the ideal law on patents for an invention should look like]." Yet, we can agree that the patent system aims overall to improve a nation's economic performance and contribute to social welfare.

Vaver offers a guiding framework that patent laws should seek to reward the right activity, on the right conditions, to the right extent, with the right level of enforcement, and the right person. He suggests that patents should more precisely match and reward the advance the inventor discloses, not be granted for activities that need no stimulus or are already adequately stimulated by other intellectual property laws, disclose all the inventor knows about the invention to as wide an audience as possible, catch only activities the patentee and the public fairly expect to be included with the patent's claims, and be enforced in ways that do not unfairly benefit patentees and unnecessarily restrain industry.[ii] Among these guidelines, proper disclosure by the inventor to the public is central.

The disclosure requirement incentivizes innovation: it prevents wasteful duplication of research, allows follow-on innovation by informing patentees about the boundaries of a patented invention, and benefits the public by disseminating new technology. In the context of AI, it may be difficult to describe how the algorithm works and to properly disclose the inner workings of the invention, putting the public in the same position as the machine. On the other hand, if AI-invented patent applications are unlikely to succeed (as is currently the case in various jurisdictions), there is a risk that inventors may choose not to disclose at all and to rely on trade secrets instead, effectively undermining the core rationale of the "patent bargain." Indeed a trade secret mandates the very opposite of patents, "locking up" the invention and ensuring no one knows about how it was made.

In assessing the various policy implications and unintended consequences of ascribing AI inventorship and ownership, we go back to trying to sort out the desired patent system. It is typical to lean on theories underpinning IP and patent law. Among the strongest and most widely adopted justification for IP protection is the Economic theory or incentive theory. In a nutshell, encouraging individual effort by personal gain is the most effective way to advance public welfare and granting rights flowing from these acts as an incentive. Although AI cannot be motivated to invent by the prospect of a patent, we can incentivize developers to develop creative machines. And having more creative machines and their patented outputs is good for society.

Natural law or labour theory is often adopted in IP, especially in copyright law. Drawing from John Locke, a person is said to hold a natural right to the fruits of their labour. So, if you put work into something, you deserve to get something out of it. According to this theory, ownership of AI-generated inventions should be awarded to the person who laboured to plant the seed, AI being the tree that bears fruit in this metaphor.

Personality theory draws from Kant and Hegel to say that a person's personality becomes fused with their inventions, thus warranting protection. Incorporating AI complicates the infusion of personality, but to what extent? Perhaps it could be argued that the developer's personality becomes fused with an inventing machine.

The theories which underpin patent law – economic theory, natural law theory, personality theory – as well as the still useful but perhaps less popular theories – contract theory, Foucauldian, feminist, Marxist, and so on – should help guide determinations of who qualifies as inventor or owner. A human touch is necessary to ensure the patent bargain continues to function. We must also ensure we are rewarding the right humans. Perhaps that means those who can be incentivized to invent to advance the public good, those who put work into the invention, or those whose personalities are

reflected in the AI or the final product. It is also essential that whichever human is granted patent protection have the willingness and ability to properly meet their disclosure obligation to fulfill their end of the bargain.

The AI Inventor v the AI Owner

The inventor is the person or persons who conceived of the invention and who is responsible for the inventive concept. It has been argued that AI should be excluded from being recognized as an inventor because modern AI processing is distinct from the human mental act of conception. Corporations have been barred from inventorship status for a similar reason: "people conceive, not companies."[iii] It has been argued in support of awarding computers inventorship that AI may function independently, and it is only sometimes the case that substantial insight is needed to identify and understand a computation invention.

In 2019, a representative of the European Patent Office stated, "It is a global consensus that an inventor can only be a person who contributes to the invention's conception in the form of devising an idea or a plan in the mind." She added that, "The current state of technological development suggests that, for the foreseeable future, AI is… a tool used by a human inventor… Any change… [would] have implications reaching far beyond patent law, i.e., to authors' rights under copyright laws, civil liability, and data protection." The implications for patent law and beyond, however, are more closely tied to who is awarded ownership than who is awarded the title of inventor.

All computer work is appropriated. Computers cannot be incentivized. The effectiveness of an intellectual property regime in incentivizing innovation and creation rests solely on its ability to impact humans and, therefore, its ability to allow the right humans to reap the rewards of AI-generated inventions or works. Further, it is human owners who will play a role in ensuring the existing model of tort liability remains functional in the context of AI-generated inventions.[iv]

Just as important as determining who is an inventor (or author, in the copyright context) – whether a human or a machine – who is the owner, and how and why, is vital. The AI patent or copyright owner will be instrumental in giving life to the right. The owner defines where, how, and when commercialization occurs. The owner enjoys the rights and incentives associated with patent protection. Ultimately the owner or owners and many other commercialization stakeholders (i.e., investors) are material to the innovation ecosystem and part of the realization of the patent bargain's public benefit.

AI's Challenges to the Patent System

The ownership considerations of the questions raised by AI inventors should not be overlooked as we work to bolster our AI innovation ecosystem for the global public good. Will the owners of the most powerful AI or those with the means to purchase the most powerful computers become patent monopolists? How can we ensure the right individuals are rewarded and have access to file a patent application in the first place? How can we not only incentivize AI innovation but also ensure this innovation takes place in a manner that fulfills the patent bargain? Which model of ascribing patent ownership concerning AI-generated inventions is put in place is closely tied to these questions.

The challenges to the patent system are complex and cannot be attributed solely to its constituent laws but also to its institutions. The entire innovation system needs to work from the pre-patent inventive stage to the post-patent commercialization stage. For instance, the speed of technological

developments can render patent examiners' knowledge dated. This is exacerbated in the AI context, where a reasonable onlooker or patent examiner may find it difficult to explain the inner workings of AI.

The peer-to-patent system piloted in the UK, Japan, the US, and Australia is one mechanism of for infusing fresh and accurate expertise into the examination process. More drastic overhauls, for example re-examining the "person having ordinary skill in the art" test, have also been suggested considering the potential high bar to obviousness that may be established as AI advances. Since data is the precondition for AI innovation, there have also been calls to ensure antitrust law and governance principles of open access and data sharing work to prevent consolidation in the data and AI industries.

When assessing the granting of patents to AI-generated inventions, it is important to consider these challenges and the wider socio-economic and cultural framework and how it can best enable the proper application of patent laws. If the computer is designated as the first owner and grants entitlements to individuals through contract, all downstream inventions would require economically inefficient assigning or licensing. If humans are listed as owners of AI-generated inventions, scholars have asked who that human should be – is it the creator of the AI, the person who asked the curious question or happened to press a certain button that set off a chain reaction leading to the invention, or the person who discovered that the AI had generated the invention? The Legal Board of Appeal of the European Patent Office recently remarked they were not aware of any case law which would prevent the user or the owner of a device involved in an inventive activity from designating himself as an inventor under European patent law. Socio-economic and cultural considerations should factor into a determination of who qualifies as user or owner and who between has a superior claim.

Generative AI and Ownership

IP ownership questions are also material in global generative AI developments. Rather than an invention subject to a patent, these generative outputs can be subject to copyright protection. Indeed, Open AI's ChatGPT, which was downloaded more than 500,000 times in its first month and accumulated more than 30 million users in two months, can generate creative works, including poems, screenplays, stories, term papers, reports, business plans and so on.

Kristina Kashtanova, the author of a comic book containing some images generated by AI, was denied copyright protection by the United States Copyright Office over the specific images generated by the AI as she was "not the author for copyright purposes." In another instance of an AI-Generated Work, Dr. Thaler filed to register for copyright protection and is now in the process of challenging a refusal.[v] While the AI was named as the author, Dr. Thaler sought ownership of the copyright. Dr. Thaler rooted his entitlement to the work in common law principles of property ownership, including accession and first possession,[vi] as well as in the work-for-hire doctrine.[vii] The outcome of Dr. Thaler's challenge will have wide-reaching implications, especially as AI-generating machines grow more popular and more advanced. Who owns the generative AI matters for determining commercialization questions and developments of various industries.

Concluding Remarks – The Hack

When attempting to harness the power of generative AI, it is key to develop skillful prompts. The appropriateness of a command impacts the quality and accuracy of the generated output. Similarly,

policy questions of great importance, such as how or how not to account for AI-generated inventions and works in existing intellectual property regimes, can be better answered when the question is more pointedly tailored. This paper serves to better prompt the discussion surrounding the AI Inventorship Challenge: it is not so much who invents or creates but who owns the patent or copyright that has the power to ensure the public benefits and the patent or copyright bargain is realized.

7. Global Governance of AI Songwriting

Patrycja Rozbicka, Simon Barber, Nicholas Gebhardt and Craig Hamilton

Over the last few decades, artificial intelligence has begun to profoundly reshape our musical experiences (Miranda, 2021; Behr, Negus and Street 2018). Technologies designed to assist musicians in creating songs have proliferated, while streaming services have transformed our listening habits through their use of algorithms (Hamilton, 2019). While AI has been used within music-making since the 1960s, we still lack an adequate understanding of its social and cultural impact. This is particularly the case when it comes to the governance structures involved in AI song production, from frameworks for regulating data to copyright issues. These are crucial questions given the size and influence of the global music industry, which generated US$26.2 billion in 2022. Below, we zoom in on a number of explanatory examples and pinpoint a need for more robust governance mechanisms to face challenges accompanying AI-assisted songwriting.

How AI is transforming songwriting

One change accompanying the rise of AI in the music industry relates to the intent of AI-assisted, as opposed to human-only, songwriting. Consider the following three examples. Incubation activity within Abbey Road Red, the innovation department of Abbey Road Studios, has led to a high level of investment in the growth of AI music companies like Humtap, Lickd, Vochlea and Lifescore. Projects like DeepMind's Wavenet (2016), MuseGAN (2017) and MuseNet (2019) have been built on music generation in various musical genres and styles using subsets of AI activity such as natural language processing, neural networks, and machine learning. OpenAI's Jukebox, a neural net that generates vocal performances, has produced an alternative genre devoted to uncanny 'deep fakes' of famous performers like Frank Sinatra (Robertson, 2020). Despite a strong sense of collaboration between artists and AI, the ultimate aim of these technologies is to create 'a hyper-realistic and expressive voice that is not distinguishable from real humans' (Stassen 2021). This raises questions about the integrity of existing bodies of musical work as AI uses these as data to improve its own songwriting abilities and generate new content.

Scholars have also begun to examine how AI is transforming musical creativity. MIMIC (Musically Intelligent Machines Interacting Creatively, 2018-2021) created a space to share creative javascript projects that use machine learning. Shimon, a musical robot created by researchers at the Georgia Tech Center for Music Technology, was fed a dataset of 50,000 lyrics covering genres including rock, hip-hop and jazz. The songs produced were composed by humans and robots together. Such advances highlight the constructive capacity of AI but potentially transform how humans can best contribute to collaborative songwriting processes.

These questions are becoming more pressing as major companies such as IBM, Sony, Google and Spotify have established labs to experiment with AI music creation. François Pachet, who leads the AI research arm at Spotify, oversaw the development of Sony's 'Flow Machines', using them to generate material in the style of The Beatles. He invited songwriter Benoît Carré (aka SKYGGE) to complete the songs, culminating in the 2018 release of Hello World, which was 'the first multi-artist commercial album created using artificial intelligence' (Nazim, 2018).

Societal and governance implications

While changes to songwriting processes and creativity or the technological investment decisions of media companies may seem far removed from social and governance processes, these transformations are in fact raising several important issues related to societal and governance challenges. For instance, the examples above open a Pandora's box when thinking about copyright law. Copyright law is one of the cornerstones of business activity in the music industries (Drott, 2020) and rests on the right of individual(s) to be recognised as creators. It is not currently clear how the music generated with AI systems will impact the legal framework within which creators currently operate, and especially the organisations which control and administer copyright. Who will own the work in question? How will creators establish copyright? And what happens when the relationships between different parties break down?

Further, as major companies get more involved in AI-facilitated songwriting, critics worry that these decisions may reinforce existing social biases while doing little to enhance accountability for socio-cultural outcomes. Debates around Google's treatment of AI researchers (Simonite, 2021), as well as reports of inherent biases in its Vision AI architecture, for example, have highlighted many of the tensions within the AI community asking 'Who Is Making Sure the A.I. Machines Aren't Racist?' (Metz, 2021). The representative bodies for marginalised groups such as Queer in AI, Black in AI and Widening NLP have taken steps to publicly denounce funding opportunities from Google. In 2022, Capitol Music Group severed ties with popular AI rapper, FN Meka, after complaints from the black community about how the musician represented "gross stereotypes" and used offensive language on recordings (Cain, 2022).

These implications are all the more concerning for their potential to exaggerate, or at a minimum perpetuate, general and pre-existing social biases in AI research and technologies. A 2019 Nesta report on Gender Diversity in AI Research noted that only 14 per cent of authors are women, and in the UK just 27 per cent of AI papers have at least one female co-author. A lack of racial diversity is similarly apparent. This has implications for how we understand the usefulness of AI technologies as well as how they are applied. A key challenge for AI developers and governance, therefore, is understanding how to address historical biases and inequities which human-technological collaboration – intentionally or unintentionally - preserves.

What next?

In this context, examining how governance can contribute to better outcomes for society and the music industry seems sensible. Given the global character of the creative industries and a high-level investment from audio streaming services and companies (e.g. Spotify, Google), we recommend a double-tier approach. First, we need a global governance solution that provides unified standards for AI songwriting (including for example, copyright regulation). Second, we should look into self-governance by companies and large conglomerates based on voluntary agreements and the creation of 'watch dogs' and initiatives to monitor corporate behaviour in the field.

Within academia, these issues are being explored through initiatives like Just AI, a network of researchers at the Ada Lovelace Institute, who are developing ethical policies and best practices in AI, particularly regarding concerns about privacy, algorithmic bias, fairness, trust and transparency. In the music industries, despite the backdrop of the recent DCMS enquiry into the economics of music streaming, there are continuous challenges to the fair remuneration of music creators and

rights holders. Multinational music corporations like Universal Music Group devote much energy to blocking 'infringing' content, including working with Spotify and Apple to police the emergence of AI-generated songs cloning the sounds of popular artists (Nicolaou, 2023). However, the recent controversy caused by the release, and subsequent takedown, of the song 'heart on my sleeve' by Ghostwriter, which provided an AI illusion of a new collaboration between Drake and The Weeknd, demonstrated that important legal and creative questions may remain unanswered for some time to come.

Dr. Patrycja Rozbicka is a Senior Lecturer in Politics and International Relations at Aston University, Birmingham, UK. Her research focuses on regulation around live music industry, interest groups and lobbying. She is a coordinator of the Live Music Project (LMP) and a head of the Birmingham Live Music Project (BLMP).

Dr. Simon Barber is a Senior Research Fellow at Birmingham City University. He leads the Songwriting Studies Research Network and is the co-producer and co-presenter of the popular Sodajerker songwriting podcast.

Nicholas Gebhardt is Professor of Jazz and Popular Music Studies at Birmingham City University and the author of Vaudeville Melodies: Popular Musicians and Mass Entertainment in American Culture, 1870-1929 (University of Chicago Press, 2017).

Dr. Craig Hamilton was Research Fellow in the Birmingham Centre for Media and Cultural Research (BCMCR) at Birmingham City University, UK. His research explores the role of digital, data and 13 internet technologies in the business and cultural environments of popular music.

8. AI Governance in Practice: A View from the Non-Profit Sector

Laura C. Mahrenbach

Much has been made about the potential of AI to enhance social good, and commentators have suggested that non-profits, as proponents of social good, may reap benefits as well. For instance, machine learning could help enhance fundraising or automate repetitive tasks, thereby freeing up resources and personnel for other tasks. But how does this play out in practice?

Claudia notes that a crucial starting point for understanding the trajectory of AI in non-profits is to clarify what kinds of data analysis are needed to achieve their strategic goals. "It's sometimes machine learning, sometimes more sophisticated statistical methods or data science, but self-learning, imitating human behavior? I would be very hard pressed to point to non-profit organizations I have worked with" these advanced AI technologies.

In fact, the use of AI technologies in non-profits is very much an ongoing process. Claudia noted that, in the initial phases, there was a dichotomous view of these technologies. Some actors, drawing on the hype about AI as a tool and looking towards tech companies and the private sector, "believed that there's outsized potential to realize opportunities and impact" by using AI. Others, often involved in advocacy or with concerns about surveillance capitalism, were more hesitant, worrying about data rights and privacy violations that would accompany these technologies. These "camps weren't talking to one another in the beginning" but have now "become more realistic and less entrenched," focused on how you can use AI while "adhering to high standards of ethical and responsible behavior."

Challenges of AI in practice

Nonetheless, Claudia says her conversations with over 400 partners working around the world over the past 6 years have revealed that, despite evolving understandings of AI and its social uses, many non-profits continue to face significant challenges in employing these technologies. Sometimes there are issues of organizational culture as employees question whether these technologies are the "best way how we can help people." Other challenges relate to problems of implementation. "Many [non-profits are very early in their digital journey. They don't have a digital strategy. They have to develop their vision of data use across the organization. They, especially the smaller ones, have very few resources, whether financial or talent."

Funding comes from a variety of sources. For instance, some foundations provide non-profits with funding because they're focused on technological issues, others because a project could benefit from AI. These philanthropic pots of money "are often small…but tend to be the most flexible" from a use perspective. Public agencies offer larger grants. However, these "might require more of an administrative effort on behalf of the nonprofits and processes might take more time." Alternatively, they may be too large for non-profits unless working through a collective. The private sector offers resources as well, often as part of corporate social responsibility initiatives. These are diverse, including money, products, software licenses – resources which Claudia describes as "very common and non-negligible." Where the money comes from can affect how AI is used in the short term, the extent to which human capital is developed, and strategic implications for employing AI in future projects.

Other challenges arise from the approach a non-profit takes when choosing a vendor as well as the conditions under which help is received and implemented. Like other actors, non-profits face strong pressure to choose a vendor based on price. As a result, the non-profit might "end up with a solution that is not the best for their problem" because that offer was accompanied by the best conditions. Relevant conditions include whether help comes in the form of a donation as well as whether "in-kind implementation help" is part of the deal. Moreover, in some instances, non-profits may "hire a consultant to manage a project because they might not have funding to bring someone on full-time." This runs the risk of "losing that knowledge once the consultant leaves." Even if they do develop sufficient technological skills during the project, "they have to think about how to deal with now being locked into a specific technology and the efforts it might take to migrate the solution to another vendor or product line" once project funding and license donations end.

Data security is another challenge looming large for non-profits using AI and digital technologies, albeit one not unique to non-profits. As Claudia noted, credit rating agencies also have data breaches, and we hear about it on the news for a few days after the leak. But such leaks are likely to have more permanent consequences for non-profits, both due to the nature of non-profits' data and their financial dependence on others for their work. "When [a data breach] happens to a non-profit, it's almost a permanent stain on there on their track record as these organizations deal with vulnerable populations and donor expectations are high." Moreover, it can negatively affect multiple funding sources. "We have seen small individual donors as well as foundations adjust funding after data breaches (or in response to questions around data use)."

Finally, social and technological contexts affect how non-profits engage with AI and digital technologies. Claudia noted that, while in technologically more advanced countries in Africa like Nigeria, South Africa or Kenya "there's more activity evolving [regarding how non-profits can use technology], there are a lot of countries where even internet access continues to still be a problem in rural areas." Non-profits around the world seek to "adopt solutions developed in the corporate sector, for example, in health" but they must do so realizing that using them in some contexts will be more difficult than others. For instance, in India, with its wealth of technological talent, there are global players such as the AAPTI Institute who work on data governance issues alongside the Open Data Institute (ODI) in the UK and many others. Yet bringing these "world-class people and world-class efforts to a farmers' collaborative or women's collaborative in rural India" remains a challenge. Importantly, this is also true in the United States, where Claudia noted organizations like Pecan Street also seek to provide farmers with data products. Nonetheless, a relative lack of technological skill and infrastructure may make such challenges potentially more difficult to solve in developing countries than in other parts of the world.

What role for regulation?

The immensity of the opportunities and hurdles facing non-profits interested in using AI raises the question of how and if regulatory bodies and governments can play a role in facilitating the former and mediating the latter. A traditional approach toward regulation would see governments setting limits and creating incentives to encourage non-profits to use these technologies to amplify their impact. In fact, regulations like the EU's proposed AI Act and its General Data Protection Regulation as well as (sub-)national spin-offs like the California Privacy Rights Act "are on non-profits' minds […] and are part of their compliance work." Similarly, we have seen initiatives by organizations like the UNICEF Manifesto on Good Governance of Children's Data addressing data

privacy issues related to children. As Claudia notes, "those things have become quasi standards in the sector and others are definitely looking to that and applying that" in their work.

Yet non-profits are also taking different paths to influence government regulations and are setting their own standards through bottom-up processes. Larger non-profits, such as Germany's Arbeiterwohlfahrt, are engaging deeply with issues of data use, considering both "how can they use these tools to increase their [social] impact, but also how can they play a role in the societal discourse around data." In other settings, we see so-called "implementation agencies" going to countries in Africa without a fully developed data strategy and "implementing digital health solutions on the ground on behalf of the government. But they are doing it in this space where there is no GDPR, there is no kind of data privacy act and similar things. So, in a way these non-profits are setting quasi standards from the bottom up."

Nonetheless, it's important to note that, for non-profits, AI is "an instrument and not the solution for every problem. We need many different instruments, and data is just one of them. But it has enormous potential to change how we operate." How? Claudia notes that much non-profit work "still treats a lot of people the same way" regardless of the social, medical or other challenges they are facing. "If we have more data, we can understand what's working in certain contexts and work with people to deliver better solutions for them."

Claudia Juech leads program development and program design for the Government Innovation portfolio at Bloomberg Philanthropies. As a former Senior Executive at the Rockefeller Foundation, the founding CEO of the Cloudera Foundation, and most recently, the VP of Data and Society at the Patrick J. McGovern Foundation (PJMF), Claudia has more than 15 years of experience leading programs and operations across geographies. At Cloudera and PJMF, she implemented a new foundation model designed to deliver the key elements nonprofits need to use data effectively: appropriate funding in combination with direct access to technology, data expertise, and long-term technical support.

Laura Mahrenbach is an adjunct professor at the School of Social Sciences and Technology at the Technical University of Munich. Her research explores how global power shifts and technology interact, with a special focus on the implications of this interaction for the countries of the Global South as well as related governance dilemmas at the national and global levels. More information available at www.mahrenbach.com. This work was supported by Deutsche Forschungsgemeinschaft (Grant No. 3698966954).

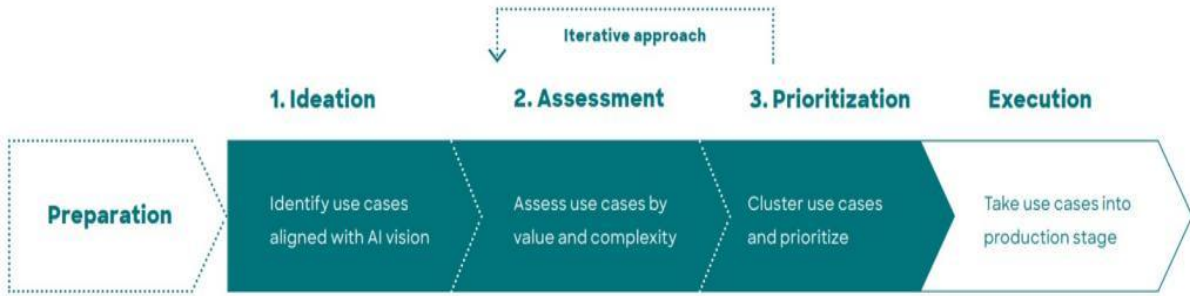## 9. AI Regulation in Practice: A View from the Private Sector

Laura C. Mahrenbach

The potential for artificial intelligence (AI) to improve business outcomes has been much lauded, particularly since the release of ChatGPT in November 2022and the subsequent proliferation of competitors. Promised benefits include enhanced flexibility, improved communication, greater productivity, and even enhanced customer satisfaction. Yet how does this work in practice? What challenges do businesses face in turning to AI? And how do political factors affect their capacity to achieve promised benefits?

I recently had the pleasure of speaking with Till Klein of appliedAI about these and related issues. appliedAI was created in 2017. Till notes that, at that time, it "was becoming really clear that there were new technological solutions to existing problems that were simply better than what we could access prior to that." AI was increasingly being viewed as "a strategic asset of nations:" businesses use it to "improve productivity on a large scale" and this could increase prosperity for the nation. Yet it was also clear that Germany and Europe were far behind the advances taking place elsewhere in the world. So appliedAI set out to "help companies and other practitioners make use of AI for their own business" and in the process also maybe "expand Germany's and the EU's share of the pie."

Challenges of AI in practice

The potentially transformative nature of AI at the business-level has led to a proliferation of advice regarding how businesses can optimize their gains from this technology. Till describes the decision to use AI as a journey. The first step of that journey is to commit to this "serious strategic initiative [which] is not going to have returns on the first day" and which requires substantial effort to realize. Many companies begin as "experimenters," trying out a single application to see how it performs and how it affects their business. As their confidence and experiences grow, they become practitioners, systematically analyzing how use cases interact and how this "creates new synergies" which are advantageous for their business. Practitioners develop AI strategies addressing their individual business needs, for instance, by creating "systematic education programs for people working at the company to build up skills" or building a "data strategy" to ensure access to good data for AI training purposes. Companies continuing along this path ultimately become very professional in their use of AI, adopting "almost an AI-first approach" where they respond to new challenges by asking "how can we use AI for that?"

While this seems straightforward, Till underlines that there are several hurdles which companies tend to face when proceeding along this path. A first challenge involves determining the use case that is most likely to yield benefits in a given setting. This is a "non-trivial task" as businesses may only invest in one use case and want to be sure they're "putting their eggs into the right basket." appliedAI suggests a framework for thinking about use cases based on ideation, assessment, prioritization and execution. In the ideation phase, businesses ask questions like, "should I focus on internal processes like production and finance and logistics or should I be focusing on the external facing activities like customer touchpoint, customer interaction, new offerings?" These ideas are then assessed via internal processes or in collaboration with partners like appliedAI. Once a business has determined its priorities in using AI, it then moves into the execution phase to begin reaping rewards from AI.

Iterative approach

| 1. Ideation | 2. Assessment | 3. Prioritization | Execution |

Preparation

Identify use cases aligned with AI vision

Assess use cases by value and complexity

Cluster use cases and prioritize

Take use cases into production stage

(Brakemeier et al 2023)

Another set of challenges arises at this stage. Till notes common ones include evaluating and enhancing staff capacity to implement AI solutions and making structural or strategic shifts which may be necessary for successful implementation. For instance, "Is AI an IT topic? Should we create a dedicated AI unit?" These questions are often difficult to answer, even for digitally advanced and/or larger firms. For one thing, incorporating technology in business activities, which companies have done for many years, "is different to developing AI solutions." To effectively use AI, you may need "to rethink your development pipeline and the skills of the people involved," which can have "implications for commercial goals and strategies" as well. Moreover, Till notes that studies have shown that "building up AI capability is one of the main entry barriers for using AI." The AI talent pool is a "very competitive market with very high salary packages" with "large companies in Germany, like the automotive ones, competing against the big tech players from overseas who have offices in Europe as well." This limits the capacity of many firms to address AI skill deficits via recruitment. Yet the more affordable option of relying on collaboration with external vendors has its costs as well. For instance, "if I use the I system from a vendor build a competitive edge, anyone else can buy it as well and there goes my competitive advantage."

Enter the political context

Till was quick to note that businesses do not use AI technologies in a vacuum. AI is "a technology that is locally developed and globally deployed" with "the main models coming from a very few players overseas, mostly big tech in the US and China, with this whole nation or winner-takes-all mentality." As he sees it, these firms "just had a bit of an advantage, like the Google search for instance, at some point in time, but then there's this flywheel effect: because it's better at one time, people use it and it generates more data, helping you to learn faster. Like so much better than everyone else, that it's really difficult to catch up. That creates inequalities for everyone else." These tendencies can be countered. For example, appliedAI uses targeted education programs and provides information to the general public on AI technologies and their uses to "make sure that AI is not only for those who can afford to hire the people, but also for everyone else."

Yet a supportive regulatory context matters as well. Till notes that, "by and large, AI regulations are welcome because, in the long run, they create clarity. It's like moving from the Wild West to some organized setting, establishing mutual expectations of what can and cannot be done while also shaping the ecosystem." Nonetheless, debates in Brussels tend to be "a lot of talk about big tech from overseas, what Facebook and Alibaba have done and geopolitical tensions. We should definitely consider this, but we should not overlook the local AI system, the innovation and startups coming from our universities across Europe." To that end, appliedAI and others "inform the negotiations from a practical perspective," generating insights from "people or organizations developing AI" and communicating these via white papers and studies. These communities provide "empirical insights" into what it would cost to be compliant with proposed regulations and how they will be affected by propositions like the European Commission's risk-based approach to classifying AI systems. The hope is that more useful regulations will be the result.

Here too Till underlines a need to keep in mind that implementing regulations like the proposed AI Act – much like the AI technologies they seek to regulate – is a process. "In the short term, it will increase the complexity of something that is already extremely complex. It will probably take us a

couple of years to establish dominant methods for how to be compliant and this will slow us down." Yet he notes experience suggests the end result will ultimately be better. When seat belts were introduced, "auto manufacturers had to upgrade and their developers were saying, 'Oh no, new rules, new this, new that,' but now clearly no one would ever use a car without a seat belt." He thinks this can be compared to where Europe is with AI at the moment: people are hesitant because of the turmoil but, in a few years, "they have this regulatory framework where they feel comfortable and maybe it will be more worth the investment than it previously was. And then we'll be happy and look back into this turmoil we have right now and say, it was worth the hassle to enjoy good quality" arising from businesses' investments in AI technology.

Till Klein is the Head of Trustworthy AI at the appliedAI Initiative, a leading initiative of applied AI in Europe, where he is working toward accelerating Trustworthy AI at scale. He has several years of practical experience with Regulations in High-Tech sectors through roles in Regulatory Affairs for Medical Devices, as Lead Auditor for ISO 9001, and as Head for Quality Management in a Drone Company. Till is an industrial Engineer by training and holds a Ph.D. (Business) on Technology Transfer and Collaboration Networks from the Swinburne University of Technology in Melbourne, Australia.