ANNUAL REVIEWS

Annual Review of Law and Social Science

AI and Global Governance: Modalities, Rationales, Tensions

Michael Veale,[1] Kira Matus,[2] and Robert Gorwa[3]

[1]Faculty of Laws, University College London, London, United Kingdom; email: m.veale@ucl.ac.uk

[2]Division of Public Policy, Hong Kong University of Science and Technology, Hong Kong

[3]WZB Berlin Social Science Center, Berlin, Germany

Annu. Rev. Law Soc. Sci. 2023. 19:255–75

First published as a Review in Advance on June 28, 2023

The *Annual Review of Law and Social Science* is online at lawsocsci.annualreviews.org

https://doi.org/10.1146/annurev-lawsocsci-020223-040749

Keywords

artificial intelligence, AI, algorithmic regulation, global governance, AI regulation, machine learning

Abstract

Artificial intelligence (AI) is a salient but polarizing issue of recent times. Actors around the world are engaged in building a governance regime around it. What exactly the "it" is that is being governed, how, by who, and why—these are all less clear. In this review, we attempt to shine some light on those questions, considering literature on AI, the governance of computing, and regulation and governance more broadly. We take critical stock of the different modalities of the global governance of AI that have been emerging, such as ethical councils, industry governance, contracts and licensing, standards, international agreements, and domestic legislation with extraterritorial impact. Considering these, we examine selected rationales and tensions that underpin them, drawing attention to the interests and ideas driving these different modalities. As these regimes become clearer and more stable, we urge those engaging with or studying the global governance of AI to constantly ask the important question of all global governance regimes: Who benefits?

1. INTRODUCTION

The topic of artificial intelligence (AI) occupies a significant amount of global policy attention. Some have seen it as holding economic promise; others as a business threat; and others, still, as a cause or catalyst for several salient social and environmental concerns. These hopes and anxieties have led to calls for global governance, which itself has become a polarizing topic. Critics argue it is captured by industry or completely ineffectual, whereas proponents and closely engaged players see themselves creating regimes that will steer the future. In this review, we frame and examine varied global governance initiatives, and their accompanying, varying AI framings, as key sites of regulatory contestation.

We first establish what both AI and global governance mean in the context of this review. Then, we take critical stock of the different modalities of the global governance of AI that have been emerging, such as ethical councils, industry governance, contracts and licensing, standards, international agreements, and domestic legislation with extraterritorial impact. We finally move to assess selected rationales and tensions that underpin them, considering the interests and ideas driving them.

1.1. What Are We Talking About When We Talk About AI?

AI has become a loose, umbrella term increasingly saddled with hype, misdirection, and confusion. In recent years, and particularly in governance discussions, it has mainly become the new term referring to technologies that used to be referred to as data mining, big data, or machine learning—or indeed, by some, practically any piece of modern software. This is to the general distress of those who seek a specific definition of AI or emphasize approaches from outside pattern recognition and statistics, such as symbolic reasoning. In this article, we are concerned less with what AI is (or should be) than with what technologies and practices are proposed to be governed when we speak of global governance of AI.

For these purposes, the term AI is perhaps most usefully understood as a practice, an "applied science and engineering discipline" (Bryson 2020, p. 4) aiming to impart qualities humans consider intelligent into artefacts in particular software. The governance of AI as a practice must therefore also be seen through this wider lens. It should include the tools and processes used by this practice, including their availability and social and material impacts. We prefer to speak of AI models, development tool kits, frameworks, data sets, or other artefacts rather than speak of how to govern "an AI," which can entail misleading connotations. This resonates with how even technical AI practitioners do not agree on the location of the "algorithm" as artefact, let alone wider disciplines whose work is important to the development and steering of these technologies (Seaver 2017). AI further includes AI practitioners and the organizations they work within; the ends to which these technologies are put; and the social, economic, and political structures surrounding their use. It also should include the characteristics of the AI artefacts themselves and how these manifest in various contexts and can affect the people, environments, and institutions they come into contact with. This review focuses on existing and clearly emerging, not hypothetical, AI technologies. Highly speculative discussions about artificial general intelligence or existential risk attract attention in some niches of academia, industry, and philanthropy-supported think tanks but are out of the scope of this article.

Some of the aspects we discuss the global governance of are AI specific, whereas others AI has highlighted but did not create, overlapping with much older issues around the governance of computing, or of social sorting and categorization (e.g., Bowker & Star 1999, Gandy 1993). We have focused on anchoring discussions of these connected areas to AI, while simultaneously highlighting that some of the most important aspects of the global governance of AI might really

be aspects of the global governance of computing technologies (or lack thereof) that interest in AI has turbocharged and made more salient than ever.

1.2. What Are We Governing When We Govern AI?

What the governance of AI does or should mean, let alone the global governance of it, is far from self-evident. We first provide an overview of a wide range of modalities of governance of AI that have clear global relevance, including public, private, informal, formal, and hybrid policy instruments, ranging from industry standards to internationally developed principles and law that cut across several areas. We then consider some of the tensions they are already having to navigate, and which may be even more visible and important in the future.

One way of understanding the global governance of AI is to look at the conceptual target of regulation. We can distinguish rules by the aspect of AI in practice they seek to shape: the development, use, and infrastructures of AI.

The governance of AI development involves attempting to apply requirements during system design and maintenance to achieve a range of policy aims. These might include broader notions such as safety or cybersecurity, specific statistical goals like narrow definition of nondiscrimination, or limited carbon emissions in training. Such requirements may also attempt to provide transparency and oversight mechanisms, such as requiring audits of AI systems before their deployment or requiring their listing in a database before their sale. Such development requirements can be national rather than global, yet insofar as they regulate activity in international markets, they may have a much wider impact (Newman 2011). Designing policy objectives into technology begs the question of where these objectives come from. For example, hiring systems sold internationally have been found to embed US nondiscrimination norms, such as the four-fifths rule, which may not tackle concerns in other jurisdictions (Sánchez-Monedero et al. 2020). Systems intended to detect hate speech or terrorist content must deal with nationally specific definitions of these in law. Furthermore, some AI-related concerns are inherently global, for example, when models with specific capabilities such as text or image generation are released internationally, or where the training process involves transnationally important carbon emissions and supply chains of extractive materials and labor (including psychological impacts from data labeling), which differentially impact upon certain jurisdictions in the value chain (Matus & Veale 2022).

The governance of AI use concerns the means and ends of deploying politically, socially, and economically consequential software. In practice, this category can range widely: laws around automated decision making in international treaties such as Convention 108+; transnational regimes concerning AI use in domestic sectors, such as medical devices, policing, and intelligence; and regimes regulating often inherently transnational uses of AI, such as in content moderation, international conflict, humanitarian response, and international policing.

The governance of AI infrastructures involves policy efforts that seek to transcend the development–use dichotomy above. This lens draws attention to the influence of transnational technology stacks by vertically integrated firms, such as Apple, Google, Amazon, and Microsoft, which sell hardware, operating systems, sensing technologies, cloud computing, networking infrastructure, and even specifically trained models, all of which can be prerequisites to the development and use of certain types of AI system. This area is less developed but is increasingly drawing the attention of competition regulators, and questions about digital sovereignty are entering international negotiations in various forms. AI systems can also be seen as reshaping points of decision making and control within organizations, which connects global issues of the political economy of these technology stacks to the real choices, constraints, and influence domestic organizations experience (Balayn & Gürses 2021).

We begin with a condensed overview of salient modalities of global governance that touch upon one or more of the above areas. In the following section, we provide a critical discussion of tensions that these modalities highlight, in terms of both their theoretical goals and practical application. We argue that the global governance of AI is still very much a contested concept, one that is currently being substantially driven by industry interests, and thus deserves to be carefully unpacked and scrutinized. We conclude by outlining several reflections for policy makers and researchers working in this field to consider when engaging with these efforts.

2. THE MODALITIES OF AI'S GLOBAL GOVERNANCE

It would be futile for any article to seek to produce a current, detailed, and exhaustive list of global governance initiatives in the area of AI. The institutional landscape is changing so rapidly, with new initiatives emerging practically every day, that such a list would be outdated before it began. Nevertheless, we can identify and illustratively populate broad, ideal-type categories that show their development across multiple issue areas. As we demonstrate, these governance modalities are all highly political and increasingly important sites of political contestation between industry, various governments, international organizations, and civil society. We organize these in (rough) order of increasing institutional complexity.

2.1. Ethical Codes and Councils

A deluge of AI ethics documents, councils, and multi-stakeholder institutions has been established in recent years. Many of these have been established by large technology firms or organizations closely affiliated and linked to them: Some are internal institutions within companies whose products' reach gives them global importance, such as Microsoft's Aether Committee, Office of Responsible AI, and Responsible AI Strategy in Engineering; IBM's AI Ethics Board; and Google's short-lived external ethics committee, the Advanced Technology External Advisory Council. Other initiatives are funded by technology firms but at least nominally external. Perhaps the largest example of this is the Partnership on AI (PAI), established by Amazon, Apple, Google, Facebook, IBM, and Microsoft in 2016. Other industry-funded fora sit at the periphery of these efforts, such as the World Economic Forum or financially supported research entities such as the Future of Life Institute or the Future of Humanity Institute.

These entities typically, nominally seek to coordinate action across the industry actors developing AI systems and aim to shape future agendas of socially responsible AI governance within and across firms. In practice, however, their efforts have been at best decidedly mixed and at worst mired in controversy. For instance, the PAI initially attracted a range of civil society members, but enthusiasm for its impact has waned over the years since its founding. Prominent nongovernmental organization (NGO) Access Now resigned in 2020, stating upon its exit that it "did not find that PAI influenced or changed the attitude of member companies or encouraged them to respond to or consult with civil society on a systematic basis" (Access Now 2020). The principles established by industry ethics councils have been criticized for being vague and practically meaningless, with no enforcement or mechanisms to demonstrate compliance. Some have argued that these voluntary efforts operate—if they operate effectively at all—downstream from the business models and company cultures at the heart of many ethical concerns (Munn 2022) or focus on engineers and design choices and are effectively subsumed into corporate logics and incentives (Green 2021). As one review of several dozen AI ethics frameworks noted "substantive divergences" among these ethical principles as to how ethical principles are interpreted; why they are deemed important; what issue, domain, or actors they pertain to; and how they should be implemented, alongside the serious absence of efforts originating in Global Majority countries (Jobin et al. 2019).

Most concerning are arguments that AI ethics principles have served predominantly as an attempt to forestall regulatory action by firms (Nemitz 2018), where "'ethics' is the new 'industry self-regulation'" (Wagner 2018, p. 84) and can provide fuel for public-facing public relations campaigns. The membership of the short-lived Google AI Ethics council, for example, appeared to be influenced by a broader policy strategy seeking to win favor in American policy circles and provide an answer to concerns in the Republican party that firms like Google privileged Democratic values and viewpoints.[1] The backlash against instrumentalized ethics has been regrettable for those arguing that moral philosophy as a field has much of use to add to discussions around AI systems (Bietti 2021).

2.2. Industry Governance

Industry self-governance beyond ethical principles today is patchy but nonetheless influential. Much hinges on the bottlenecks firms have established in the AI space. A small number of firms control the resources that facilitate the training of AI systems, including physical (such as large clusters of graphical processing units connected to cheap sources of electricity), epistemic (access to cutting-edge groups of researchers, the ability to pay huge salaries to university professors), and informational (access to data sets, the resources to pay for extensive labeling, and live systems for collection and experimentation) resources. Industry actors have sponsored significant quantities of research into technical tools, particularly through high-profile academic conferences, which some scholars have characterized as a process of "manufacturing consensus" around political concepts like fairness (Young et al. 2022) and promoting shallow, decontextualized, engineer- and lab-centric methods that scale smoothly and cheaply (Gansky & McDonald 2022, Green 2021). Such industry governance can be seen as a form of transnational policy entrepreneurship, consisting of the "management and communication of expert discourse rather than the data, evidence or research findings" (Stone 2019, p. 1133).

One particular direction tangible global AI governance is fast gravitating toward is the lucrative and publicly salient domain of general-purpose systems. AI systems designed for generic forms of analysis are classic dual-use technologies, supporting both relatively benign purposes, such as stock image or boilerplate text generation, and controversial ends, such as criminal activity or democratic disruption. An important question arises—how to facilitate good purposes while limiting or preventing unwanted ones.

Many of the most performant AI systems on the market of this type are sold as-a-service through new, platform-based business models premised on cloud computing (Cobbe & Singh 2021). These models are marketed as generic, foundational capabilities in areas such as text or image analysis or generation, which users can integrate and tailor to specific applications. Although users often bring their own data sets to fine-tune the model to their use cases, they rarely have a full copy of the resulting model and are instead typically permitted to use them only as question–answer systems via application programming interfaces.

This form of distribution of AI capability gives platforms the ability to act as important governance decision makers. Insofar as desirable models are proprietary, their usage can be made conditional on certain use purposes. Google, for example, restricts their celebrity recognition facial classification system to white-listed customers in the media industry (BSR 2019). Some commentators have expressed concerns that governance becomes difficult or impossible if models are open sourced, such as the production of text without steganographic watermarks to tell it is artificial

[1]It was shuttered after a swift backlash centering on Google's appointment of the trans-exclusionary director of the Heritage Foundation, a think tank associated with the Republican Party, to the board; the company has not since replaced it with alternative mechanisms (see Bietti 2021).

(Aaronson 2022). Furthermore, the funding of the kind of labeling needed to support outputs that are not obviously harmful, such as the generation of child sexual abuse imagery and narratives, costs hundreds of thousands of dollars even for a single model, using outsourced labor paid less than $2 per hour, and with limited psychological support or counseling (Perrigo 2023).

It is currently unclear just how indispensable platforms will be in the provision of AI in the future. Multisided platforms for AI models are emerging, such as Hugging Face, which provides infrastructure for the distribution and use of systems made by others. In our current digital environments, select entities are extremely powerful regulators. App stores are the only way to install certain types of software on many mobile devices, and they regulate issues such as content and data usage of software through a mix of technological restrictions and contractual limitations (Cowls & Morley 2022, Marsden & Brown 2023, van Hoboken & Fathaigh 2021). There are constant concerns around the consistency and quality of such rules—YouTube regulates its users differently by tier (Caplan & Gillespie 2020), whereas Facebook has breached Apple's rules on several occasions in ways that would likely have seen less prominent software banned entirely (Carman 2019).

In the future, global AI governance seems likely to become highly enmeshed with platform governance. If important intermediaries remain, they will be both powerful governors and a target for legislators to regulate AI systems (Cobbe et al. 2023). This would mirror countless prior experiences in information technology law, where intermediaries in areas from social media to cryptocurrency become "regulatory access points" (Finck 2018) or "chokepoints" (Goldsmith & Wu 2006, Tusikov 2016). As it stands, the Council of Ministers of the European Union is already considering placing some obligations on providers of general-purpose AI systems in the draft AI Act, suggesting further that they can avoid certain legal obligations by prohibiting particular uses of them in good faith and acting upon the detection of "market misuse" (Counc. Eur. Union 2022a, title IA).

2.3. Contracts and Licensing

Another emerging, important, private form of transnational governance over AI systems is the use of contractual terms to attempt to limit the uses of AI and its outputs. This mechanism appears to be inspired by important innovations in the history of computing governance—in this case, the intellectual property regimes that emerged around open-source software.

In response to the proprietary, corporate-controlled nature of much AI development, and spurred at least in part by growing concerns over digital sovereignty and national dependency on certain significant AI systems and capabilities, interest in creating open-source, general-purpose AI systems has grown in recent years. Platform marketplaces of sorts, such as Hugging Face, have emerged to facilitate the querying, distribution, and development of open-source AI systems (Cobbe & Singh 2021). However, the ease of access to these systems has posed some political questions, because few checks and balances exist to prevent their application to controversial tasks. Some model designers have in response turned to contract law, retaining the intellectual property of a model (i.e., not releasing it into the public domain) and instead providing only a conditional license to use it (Contractor et al. 2022). Such licenses have long been a central part of software and digital intellectual property (IP) governance. The copyleft family of licenses, such as the GPL license, require any derivative work based on the licensed code to be distributed under the same or equivalent license terms, whereas Creative Commons licenses allow content to be reused under a menu of conditions, such as attribution or nonalteration or for noncommercial purposes (Guadamuz 2004).

In the AI space, the RAIL (Responsible AI Licenses) initiative claims to go beyond these initiatives to add "behavioural use restrictions." The IP holders applied such a license to the Stable Diffusion model, an attempt to make a model similar to OpenAI's proprietary DALL-E 2 image

generation model. This license forbids users to use the model, inter alia, to defame others; to provide medical advice; for law enforcement or similar purposes; for purposes that intend or have the effect of discrimination, broadly understood; for certain fully automated decisions, a provision based on data protection law; and to exploit individuals in a way likely to cause physical or psychological harm—rules that seem to be lifted directly from the European Commission's draft AI Act (Rombach & Esser 2022). Although these contractual terms are not perfect, and some of these provisions are vaguely drafted (for example, drawn from laws without importing the relevant definitional sections, or not specifying jurisdiction for defamation, or general illegality), the most significant barrier to their effective implementation is that only copyright holders can enforce the copyright license (Contractor et al. 2022, Guadamuz 2004). Although software companies may be fairly aware of a limited number of competing software companies in the area, and able to gather enough information to pursue effective enforcement, the prospects for successfully preventing social misuse seem less likely when the issue is around the use of technology rather than its development and purported IP theft. Relatedly, open-source or public interest teams developing AI systems licensed in this way likely would lack the ability to spot and police the huge variety of users of a general-purpose AI system, nor would they have the necessary legal resources to make such governance realistic at scale. We might envisage another institutional layer of enforcement mechanisms—just as the foundations holding copyright to much open-source software have created "community-oriented" enforcement principles (Ballhausen 2022)—but to our knowledge no such proposals have yet emerged.

A more legally tenuous effort can be seen as some platforms' attempt to license models' output, rather than the models themselves, in an effort to govern their global use. OpenAI, a firm known for its models that can generate novel text and imagery, operates a content policy requiring generated media to be disclosed as artificial and forbidding certain topics, such as that depicting "illegal activity," relating to politicians, or promoting "major conspiracies." If these guidelines are breached, OpenAI, which claims to hold ownership of generated media to the extent permitted by law, reserves the right to revoke the license it provides to users. The extent of their effective enforcement based on IP, however, is probably limited by two factors: the jurisdiction and the limited creative skill or labor required to make the prompts the systems use to generate the media (Guadamuz 2017, 2022). Elaborate prompts may lead to some protection, but this would seem a limited mode of governance, because there is no reliable connection between prompt complexity and potential harm.

2.4. Standards

The historical global governance of computing has relied heavily on self-regulatory fora for creating engineering standards (Harcourt et al. 2020). Such standards have always been crucial for networked technologies, which require components to play by the same rules to achieve functionality. Bodies such as the Internet Engineering Task Force, the World Wide Web Consortium, and the Institute of Electrical and Electronics Engineers (IEEE) manage important standards such as TCP/IP, HTML, and 802.11 (WiFi). However, outputs of these organizations have political dimensions, going beyond facilitating functionality to produce value-laden, human rights–related design patterns and outcomes, such as those relating to privacy and free expression (Braman 2011, Cath 2021). The international proliferation of unified networking standards gives them a significant global governance role (DeNardis 2014), and substantive policy decisions and values can effectively piggyback on their functional necessity and be baked in at a technical level.

Standards processes can empower well-resourced incumbents. In the networking domain, the adoption of standards by large actors can coerce other actors to adopt the same changes (Cohen 2019, ten Oever 2021). Involvement in any standards body requires significant investments of

labor, and those with experience of the process are more able to influence it. Cost can further limit accountability. Many standards relevant to computing are made through proprietary standards bodies, where the process of standards development is closed to those who are not paid members and the final product is proprietary, often costing hundreds of dollars to obtain the entire interrelated set of rules needed.

Despite this, the success of networking standards has led some voluntary standards organizations to seek the same approach for the governance of AI systems. One of the earliest sets of standards to emerge was the IEEE P70xx series. These include published standards on transparency (7001–2021), processes for considering ethical issues in design (7000–2021), and standards on bias and "ethically-driven nudging" (7003TM, 7008TM). The International Organization for Standardization (ISO) also has a range of standards, mostly under development, through the committee ISO/IEC JTC 1/SC 42, established in 2017. Both of these organizations typically work on a subscription model; retain copyright to the standards published; and make money either through licensing access directly or through licensing localizations and translations from typically private national standards bodies, such as the American National Standards Institute (known as ANSI), the British Standards Institute (known as BSI), or the German Institute for Standardization (known as the DIN). National measurement institutes, which unlike most standards bodies are typically public entities, have also been working on AI standardization, including the National Institute of Standards and Technology in the United States, which has created an AI Risk Management Framework (NIST 2023), and the National Physical Laboratory in the United Kingdom, which is a partner in the United Kingdom's AI Standards Hub.

Not all standards are necessary for functionality. Some might simply transmit knowledge to lower-capacity organizations looking to understand how to build a system in line with current best practice, resembling a simple knowledge output instead of a coordination mechanism. Another important role is signaling. Computing standards can signal best practices to other market actors, such as to actuaries for insurance purposes (Shackelford et al. 2015, p. 352). They might be used as a signal for legislators when they are considering whether firmer regulation is required (Marsden 2011, pp. 57–58) or to courts when assessing tortious negligence (Schepel 2005, chapter 10).

Governments appear to be taking two broad approaches to AI standard setting, which can be classified, following the standards and certification literature, as hybrid or symbiotic (Cashore et al. 2012). Hybrid approaches seek to incentivize private standards to allow compliance with a legal instrument, as proposed in the EU AI Act (discussed further below). Symbiotic approaches see private certification systems reinforcing the authority and legitimacy of other governance systems; an example is the optional industry certification mechanisms that are in EU data protection and cybersecurity law (Kamara 2017). As standardization incentives go beyond functionality and network conformity, the regulatory trajectory of AI standards seemingly resembles non-computing standardization domains, such as sustainable products or product safety, more than the typical networking standards (Matus & Veale 2022).

2.5. International Agreements

In parallel to international industry self-regulation, a series of intergovernmental standards and fora around AI issues have emerged. Standards and principles include the Organisation for Economic Co-operation and Development (OECD) Recommendation on AI (OECD 2019), the UNESCO Recommendation on the Ethics of Artificial Intelligence (UNESCO 2021), and the G20 AI Principles (G20 2019). For the most part, these do not differ significantly from industry documents or, arguably, industry interests. For example, a synthesis of principles did not surface recurring mentions of digital competition, power, or control over technologies (Cowls & Floridi 2019). Despite its global policy salience, digital power appears to be an issue technology firms are

keen to keep off the table. A range of countries have set up a Global Partnership on AI, not to be confused with the private sector–led PAI discussed above, to build on the OECD work, although as it stands its impact is unclear.

The Council of Europe (CoE) has taken a leading role in AI governance, as it has regarding previous digital and data-related legal instruments. The 46-member club is most notable for the European Convention on Human Rights, which underpins the European Court of Human Rights and much international surveillance and profiling jurisprudence. It also spearheaded the Budapest Convention on Cybercrime; the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data; Convention 108 (and more recently 108+), which underpins international data protection standards; and the Oviedo Convention, which attempts to safeguard human dignity in the context of more sophisticated biotechnological innovation. Despite its name, the CoE's treaties can be global in influence: The Budapest Convention on Cybercrime has 67 ratifications or accessions, including many countries beyond Europe, such as Canada, the United States, Australia, Nigeria, and Japan.

In September 2019, the CoE set up an intergovernmental committee, the Ad Hoc Committee on AI (CAHAI), to examine the feasibility of a horizontal and transversal legal framework for the development, design, and application of AI systems. In June 2022, the Committee of Ministers instructed the successor body to the CAHAI, the Committee on Artificial Intelligence, to "proceed speedily with the elaboration" of such an instrument, following CAHAI's work on the possible elements of a treaty. The proposal from CAHAI diverges clearly from the principles above in recommending provisions around effective compliance and independent national supervisory authorities, and in recommending the prohibition of certain types of AI systems, such as those engaging in biometric categorization or social scoring. At the time of writing, the Committee on Artificial Intelligence agreed at the second Plenary Session, at the request of the United States and to significant criticism from civil society bodies who claim no such final decision was made, to draft iterations of the proposed Convention in a closed drafting group that consists only of potential Parties to the treaty. Media reports highlight that this was under pressure from the United States, whom the CoE has significant diplomatic interest in attracting as a signatory, which did not wish its negotiating positions on the narrowing of the Convention, such as its exclusion of private sector bodies in their entirety, to become public knowledge (Bertuzzi 2023). Simultaneously, the European Commission managed to obtain the authority to negotiate on behalf of EU member states, concerned that "the convention may affect existing and foreseeable future common Union rules or alter their scope" (Counc. Eur. Union 2022b).

2.6. Converging and Extraterritorial Domestic Regulation

The most concrete form of AI governance involves rules set by national governments, or by supranational associations of national governments like the European Union. In areas of digital policy as in other areas, these regulations often have a transnational impact, because they not only affect international businesses seeking to offer products in the market in question but also—under certain conditions—create incentives for firms to harmonize their rules to reduce costs across jurisdictions, potentially adopting new national regulations as a de facto international standard (Bradford 2020, Kalyanpur & Newman 2019).

The European Union and its member states have quickly emerged as a key player in the global regulation of AI, passing and proposing a variety of new regulatory instruments since 2016 with impacts on machine learning, in addition to the application of existing, generic instruments. Most recently, the European Commission has proposed a draft AI Act, with a primary aim of setting harmonized, maximum standards for the characteristics of AI systems in certain high-risk application domains, such as educational assessment, hiring, or aspects of policing and the

judiciary (Veale & Zuiderveen Borgesius 2021). Providers of these systems would be responsible for self-certifying that they meet certain essential requirements, which would be elaborated in proprietary standards authored by private European standardization bodies, European Committee for Standardization (CEN) and the European Committee for Electrotechnical Standardization (CENELEC), and their national members. Any providers wishing to sell high-risk systems into the European market, or to deliver high-risk AI systems from a distance, must certify their systems in accordance with European standards. As a result, providers wishing to deploy a model in markets across the world are likely to have to pay attention to these standards, which as proposed include considerations of issues such as bias, human oversight, and cybersecurity, and which come with certain public documentation requirements.

Beyond the draft AI Act, many other European regulatory instruments have the effect of regulating some aspect of the use, development, or infrastructures of AI, with varying global dimensions. The proposed Platform Work Directive is another example, as it advocates for aspects of transparency and oversight of systems used in algorithmic management (although its territorial scope is limited to where the platform work is actually performed, meaning extraterritorial impact is limited if international platforms choose not to use EU-based workers). Both the Digital Services Act [Regulation (EU) 2022/2065, DSA] and the Fairness in Platform-to-Business Regulation [Regulation (EU) 2019/1150] involve some provisions relating to the transparency of AI-powered recommender systems in the context of significant online platforms offering select services (Cobbe & Singh 2019). Because recommender systems tend to span borders, this transparency will come with global impact. The debates around the extent to which platforms deploying recommender systems are liable for content that they recommend, and the manner in which they recommend it, have also created important legal questions that are being explored by courts [e.g., Case C-682/18, *YouTube and Cyando* ECLI:EU:C:2021:503 (EU Court of Justice) and case *Gonzalez v. Google LLC*, 598 US __ (2023)]. These have particularly global importance because they concern the extent to which the use of AI can be covered by wide-ranging intermediary liability shields, which can preclude other forms of governance.

Alongside these AI- or platform-specific forms of regulation, there are many areas in which established areas of government policy making have implications for AI systems. Data protection law is one example. Data protection law around the world is often similar in form, with many instruments both in the European Union and beyond containing similar provisions to Article 22 of the General Data Protection Regulation [Regulation (EU) 2016/679, GDPR] (Binns & Veale 2021, Demetzou et al. 2023), an article that provides for safeguards and legal bases necessary for significant decisions and measures taken solely on the basis of automated processing. Article 15, which provides for access to the logic of the processing of some significant profiling systems, similarly has been implemented across numerous jurisdictions. These provisions represent some shared minimal baseline of information and, in some cases, transparency regarding significant uses of AI without human involvement but are drafted in a way that makes them difficult to apply in many realistic situations (Binns & Veale 2021, Custers & Heijne 2022). Statutory restrictions in data protection law with implications for AI development, as opposed to AI use, are rarer but exist. It is generally illegal to train a machine learning system on personal data the developer cannot establish a legal obligation to process, such as data from a leak or hack. Insofar as AI systems, particularly large language models or other generative tools, can be considered as personal data sets, this may restrict their movement and sale, which may have cross-border implications (Veale et al. 2018).

Intellectual property law interacts with both domestic regimes and other regimes of governance discussed above. Copyright law may prohibit use of certain personal and nonpersonal data without appropriate authorization. The most relevant global form of such authorization is the

text and data mining exemption, versions of which were first implemented in Japan in 2009 [now Copyright Act of Japan, articles 30–4, 47–5], followed by the United Kingdom in 2014 [Copyright, Designs and Patents Act 1988 s. 29A], and later the European Union in its Digital Single Market Directive (EU Directive 2019/790) (Caspers et al. 2017). These exemptions are becoming increasingly relevant as unrest grows over scraping and use of text and multimedia from the Web, particularly from skilled artists, writers, and coders, for use in the training of large-scale generative models. Whether individual users of such models are granted rights over the outputs of models they use, or whether these outputs do not qualify for protection such as copyright, also affects the scope of the contractual global governance outlined above. Furthermore, the international contours of trade secret law also serve to limit the scrutiny of AI systems and data sets by regulators, and the transparency that can be mandated of deploying organizations (Radauer et al. 2022).

In certain cases, competition law may prevent combining data across services and potentially prevent the unification of certain AI infrastructures, if those might have broader anticompetitive or injurious effects. Competition authorities are exploring the interaction of AI and their existing powers, while they and legislators also look to new ex ante instruments, such as the Digital Markets Act (EU Regul. 2022/1925, DMA) or the proposed UK Digital Markets Unit. However, little enforcement directly around AI has taken place to date, as regulators have typically considered the underlying data infrastructures rather than AI directly.

3. RATIONALES AND TENSIONS IN THE GLOBAL GOVERNANCE OF AI

While the global governance of the development, use, and infrastructures of AI is currently in a state of rapid activity, it is worth stepping back and unpacking the underlying drivers and implications of global governance efforts. Theoretically and empirically, a clearer understanding of the rationales and limitations of global systems of regulations can illuminate the potential of these efforts—but also the areas where key tensions and shortcomings are likely to emerge.

3.1. Governance for Transboundary Issues

The first, and perhaps most obvious, rationale behind global efforts is geographical—specifically, the scale and transboundary nature of these systems and their impacts. The value chains that link the supply and implementation of AI technologies, and the infrastructures on which they rely, have transboundary elements, and governing them by a single set of rules is sometimes technically advantageous for functionality—as with internet protocols—as well as economically advantageous for businesses that wish to be governed by a single set of rules.

Transboundary global governance also emerges from a concern about transboundary impacts. AI value chains have analogies in commodity supply chains (Matus & Veale 2022). This implies that the infrastructures that support data collection and training, for example, may perpetuate harm on underpaid workers in countries well beyond the labor protection regulations of the software firms making use of these data. Related harms have already been demonstrated in the context of content moderation in the platform economy (Gray & Suri 2019, Roberts 2019). In recent years, transborder litigation has emerged concerning the harms of data labeling on marginalized workers, such as from NGO Foxglove against Meta Platforms and their Nairobi-based former contractor for content moderation, Sama (Perrigo 2022). At the same time, Sama cut ties with generative AI model firm OpenAI due to the "traumatic nature" of the required data labeling (Perrigo 2023). Because such labeling practices are highly dematerialized, informationalized, and globally mobile, there is some argument that resultant governance issues must be dealt with at the global level.

Of course, global governance is not the only possible response to transboundary impacts. These can be dealt with via governance tools that range from formal, (nonglobal) trade agreements to supply chain contracting requirements. Yet these more geographically limited approaches have potential drawbacks that may drive responses into a global sphere. In this case, the well-established argument about races to the bottom provides a useful perspective. The argument, which emerged in the context of environmental regulatory scholarship, is that in the absence of a centralized transboundary governance regime, firms would be incentivized to relocate their most environmentally harmful activities to areas of low regulatory stringency, cost, and/or enforcement (Esty 1994, Konisky 2007, Porter 1999). To remain competitive, states or nations with stricter regulation would be forced to lower their standards, leading to an overall downward trend in regulatory standards. Some studies pointed out numerous reasons that firms did not, in fact, relocate operations into areas with lower standards (including costs, firm reputation, and others). However, Porter (1999) showed that poorer, industrialized countries faced this downward regulatory competitive pressure, whereas higher-income, industrialized, high-stringency countries did not. In other words, the lack of centralized global regulations means that those with weaker regimes will compete based on looser regulation, even if that has local negative impacts.

Although much of this original work was focused on environmental regulation, studies have expanded into labor and other domains of regulation—finding, for example, that decreases in labor regulations lead to increases in foreign direct investment (Olney 2013). For AI systems, the implication of concern is that some places will become havens for more harmful activities (i.e., low-wage labor) or the development of problematic software that would be regulated (or even prohibited) in other places. Given the ease with which algorithms and data can move, this could undermine attempts to control systems capable of uses already subject to proposed regulation, such as uses in areas of psychological manipulation, as offensive physical or cyber weaponry, or as surveillance systems. It may also hinder efforts to govern development or infrastructures, as firms in the AI value chain may search for jurisdictions where local environmental or labor policies create more economically competitive conditions for data servers or labeling services (Ensmenger 2018).

But it would appear that as of yet, these potentially legitimate rationales for a global governance system, to prevent havens for harmful development practices, or for the development of AI products restricted elsewhere, are not well-reflected in current efforts. In these potential framings, global governance is a state-led activity. Although some private regulatory systems have played a role, for example, around sustainable commodity certification, they are far short of replacing each other. Indeed, privately led governance systems often compete among themselves (Lambin & Thorlakson 2018, Steer. Comm. State-of-Knowledge Assess. Stand. Certif. 2012). State-led approaches reduce negative competitive impacts while also preserving political legitimacy—and political legitimacy is lacking in the current suite of AI global governance efforts (Erman & Furendal 2022).

3.2. Governance as Lobbying

Governance mechanisms themselves can also play a role in facilitating actors to obtain preferred policy choices and outcomes, both within a particular regime and also considering the interactions across regimes. We look here at both efforts to forestall harder governance approaches, such as law, and strategic signaling to amplify influence.

3.2.1. Forestalling regulation. A more cynical take on the current efforts, if they are not particularly well aligned with the above rationales, is that they are attempts to forestall exactly the kinds of state-driven governance outlined previously. The nature of the interaction between

private, industry-led voluntary governance efforts and existing, or future, state-led binding regulation has been a matter of long-standing debate for global governance scholars and transnational regulation researchers. Whether the efforts involve corporate responsibility programs (Auld et al. 2008), product certification schemes (Bartley 2010), or broader forms of company-led "regulatory standard setting" (Abbott & Snidal 2009), various voluntary commitments made by firms can be deployed strategically to undercut policy-maker and public demand for stronger, more robust rules. They can put firms in an advantageous position in which, after having declared various commitments (or set up industry self-regulatory bodies that seek to coordinate voluntary standards, practices, and norms), those commitments can simply be walked back or never effectively implemented when it is convenient (Malhotra et al. 2019).

Technologist-led efforts for governance can portray a policy area as filled with uncertainty, in need of knowledge that is just around the corner—but that will never come. Computer scientists may think of this in terms of the famous halting problem—an endeavor that may, by its very operation, never finish. Portraying AI to potential regulators as a highly technical area takes advantage of the public sector's enamor with technology firms as "emissaries of the future," deserving of epistemological deference (Zuboff 2015), and these firms' accumulation of AI expertise places them at the heart of rule making. The argument that legislators might blunder into rules that harm an industry through misunderstanding can be a politically powerful one. Yet the knowledge situation around AI governance is not one soluble by more expertise but is best classified as "postnormal" (Funtowicz & Ravetz 1993), challenging the classic role science is seen to play in problem-solving. The issues that AI governance is often truly about are not technical but deeply normative and distributive—which actors make decisions in society, who bears risks and errors, and what justice should look like procedurally and substantively (Balayn & Gürses 2021). Scientific consensus is unlikely to be achieved here where facts are uncertain, values are in dispute, and decision stakes are high—and as a result, framing the issue in terms of technological or epistemological uncertainty acts as a delay tactic. Although some dream of standardizing, encoding, and integrating values into transparent algorithms—just like applying a technical patch—this imagines we can and should lock politics into place, when new political claims and movements have been so crucial to our societies and require constant reflection on algorithmic practices in their broader contexts (Amoore 2020).

Some of the values at the heart of issues of AI governance may also be expected to differ nationally due to different local and regional political preferences, cultures, or economic situations. Forcing issues with international divergence onto the global scale creates a need for consensus where diverging national regulation may be an appropriate alternative approach, at least in the short to medium term. The politics of subsidiarity—achieving some sufficient outcome at a lower level—is fundamentally incompatible with many technology firms' chosen platform business models, which value unimpeded (or unimpedable), low-friction scaling as deeply foundational (Pfotenhauer et al. 2022).

3.2.2. Signaling. Some research has suggested that by appearing socially responsible, firms can gain closer access to policy makers (Werner 2015). The voluntary creation, funding, and support of the governance methods described above have been woven alongside corporate capture in the responsible AI research space (Young et al. 2022). Taxonomies, logics, and ways of thinking, examining, assessing, and framing are likely to be drawn into state-led policy discussions from industry research simply due to the proportion of industry research on these topics in open academic fora. This signaling of interest and expertise seemingly leads to situations where "expert groups" of public sector bodies are staffed largely by the entities likely to be subject to this regulation (Veale 2020). This is not to say that many of the individuals within these firms are not experts

in this area—it is precisely the rapid hiring and accumulation of research teams, particularly from academia, into industry labs and publishing in traditionally academic venues that have fueled the number of industry representatives in any given meeting when any AI-related governance issue is being discussed. In recent years, the tensions of holding a critical view on technology while also benefiting from an industry research position have become more apparent, as a range of individuals have been fired for organizing or publishing research within large technology firms such as Google (Whittaker 2021).

3.3. Governance as Framing

Setting up new institutions, norms, and practices is a creative act that can go in many directions. It requires reasoning about the future, considering how both technologies and economic or societal relations will develop. Those setting the stage have considerable framing power, the influence of which we now turn to consider.

3.3.1. "Talking AI into being." Scholars of science and technology studies have long been interested in how texts about the future shape the future. Envisaging governance efforts regarding future technologies involves creating expectations that in turn empower certain actors or efforts over others (van Lente & Rip 2012). "Involvement with the making of a governance model can thereby make both successful scientists and successful policy-makers" (Voß 2014, p. 323). In AI governance, national strategy documents have been explicitly highlighted as having performative dimensions—trying to "talk AI into being" (Bareis & Katzenbach 2022). Global governance can be seen in similar terms, both entrenching certain visions and their accompanying business models into the policy discussion and drawing further attention to the companies at the core of AI as those realizing humanity's technological future.

3.3.2. State-like ambitions. Parallel efforts have been seen in other areas of digital policy. Efforts to regulate content online, in an environment of higher public salience and regulatory pressure, are increasingly proactively led by industry and industry-led consortia (Caplan 2023, Gorwa 2019). One particularly notable example is the Oversight Board, an initiative created and funded by Meta, which pays a salary to 23 individuals (high-profile academics, civil society experts, and former politicians) to review a small set of the content-moderation decisions that the company has made and make policy recommendations (Douek 2019). The initiative has been polarizing, with some considering it an innovative approach to tricky transnational speech governance, whereas critics are more inclined to see it as a significantly flawed institution and possibly even a predominantly public relations–driven exercise (Arun 2022, Klonick 2019, Pallero 2020). Regardless of its direct impact, the Oversight Board has been part of a broader strategy of legitimation that benefits Meta/Facebook (Dvoskin 2022): It has allowed the firm to use "constitutional metaphors" and piggyback on the language of law and state building to further portray themselves as acceptable, trustworthy rule makers (Cowls et al. 2022). For instance, the project created a charter that begins with the grand framing of the US Constitution ("We the people. . ."), and early iterations of the initiative infamously referred to it as the "Facebook Supreme Court"—a framing that successfully shifted the discourse around the Oversight Board into a register that invited comparisons between Facebook and the legal systems of democratically founded states (Klonick 2019), rather than to classic forms of industry-led informal regulation and/or lobbying (Gorwa 2019).

In the AI space, efforts like the PAI have some similar characteristics. By creating an ostensibly multi-stakeholder but in practice largely industry-led and -controlled policy forum (its founding ten-member board featured seven industry representatives, along with one academic, one funder,

and one civil society representative),[2] powerful platform multinationals can seek to not only influence the agenda for AI's global governance, ensuring that it adheres to their broad interests, but also portray themselves publicly (and to policy makers) as being actively engaged on the deep questions of responsible and socially beneficial AI innovation. In doing so, such efforts can frame policy areas and actors in influential ways, such as in labor, considering speculative automation of jobs rather than the very real issues facing workers in the global data-labeling and -processing pipeline (Posada 2022, van Doorn & Badger 2020), or on a generic notion of "AI for Climate" rather than the growing environmental issues posed by the production, use, and disposal of information and communications technologies (Belkhir & Elmeligi 2018, Dobbe & Whittaker 2019).

3.3.3. Inappropriate AI-centricity. Seeing governance issues as characteristics of AI systems is itself a choice, and one that is not always clearly appropriate. Even the classic algorithmic issues that have occupied the research literature are in many ways not centered on the model at all. In issues such as transparency, it is not easy to separate the intransparency of an AI system from the inscrutability of its use, the knowledge and training of its users, the availability of its documentation or code, oversight over its outputs, or simply the human interpretability of its input data (Kemper & Kolkman 2018, Vaughan & Wallach 2021). Despite the metaphor of the black box, many of these boxes themselves are deeply mired in shadow, caused, for example, by the opacity of business models or public sector contractors. Even a white box looks black in darkness. Issues of discrimination cannot be easily separated from dynamic processes of data collection and retraining, or the nature of the decisions or measures linked to the AI systems' outputs (Lum & Isaac 2016). Issues of failure modes relating to particular tasks cannot be easily separated from the problem framing of those tasks, some of which may not be amenable to prediction or automation at all (Birhane 2021, Passi & Barocas 2019). Issues of accountability can be separated from neither the much longer sociotechnical process (Cobbe et al. 2021) nor the need for an active, critical audience (Kemper & Kolkman 2018).

More broadly, concerns stemming from the use of power in areas using AI may stem from technological areas largely disconnected from debates and framing in AI governance. The design space for recommender systems is heavily entwined with closed, monolithic social media platforms (Keller 2021). The use of facial recognition is closely tied to the installation and repurposing of cameras in public spaces. Proposals for machine learning to algorithmically and erratically scan and flag audiovisual media inside encrypted messaging channels (so-called client side scanning) are made possible by locked-down end devices with limited avenues for scrutiny or installation of alternative software (Abelson et al. 2021, Gorwa et al. 2020). This points to the disconnect across the global governance of technologies. It also highlights the significant risk that the global governance of AI is a chosen—and hyped—venue by potential regulatees specifically to keep deeper, more structural questions out of scope. Framing issues as technological tends to lead to framing solutions as technological (Green 2019). Technology firms are prime beneficiaries of this outcome.

3.3.4. Not quite global. A crucial question for global governance has always been who has a seat at the table, and what kind of power they have to govern. If one takes a historical perspective on global governance as relating to broader patterns of international relations and to the various interactions between states and other emerging actors seeking to "govern the globe," the exclusion of certain groups has always been arguably central to it as a project (Avant et al. 2010). One might argue that "contemporary global governance started as a set of institutions designed to help secure the global dominance of white folks by managing some of the conflicts among Europe's imperial

[2]The current Partnership on AI team is listed here: **http://web.archive.org/web/20181020062639/https://partnershiponai.org/team/**.

powers" (Murphy 2015, p. 192). For this reason, one must pay close attention to the actors being involved in the new institutions of AI governance, and the patterns of representation and exclusion being enacted in them.

Many efforts led by governments, international organizations, and industry to develop policy initiatives for global "AI for Good" or "responsible AI" nonetheless fall into a "paradox of participation," involving the surface-level participation of Global South stakeholders without providing the accompanying resources and structural reforms to allow them to be involved meaningfully (Png 2022). The infrastructures, development, and use of AI in Global Majority countries are—despite the various global ethics guidelines and multi-stakeholder governance efforts ostensibly being developed to shape them—seemingly characterized by systemic logics of material extractionism and commercial exploitation (Png 2022, p. 1438). This has led to a current dynamic where, rather than accepting depoliticized inclusion in friendly processes of allegedly win–win global governance, political actors in a diverse set of Majority world contexts have been playing an adversarial, "challenging function" to the emerging AI governance regime, actively seeking to resist and subvert it through a host of creative techniques (Png 2022, p. 1439).

4. CONCLUSION: POWER IN THE GLOBAL GOVERNANCE OF AI

If a sort of global AI governance regime—composed of an interlinked arrangement of overlapping, transnational sets of private standards and best-practice-setting initiatives, as well as normative declarations and principles promulgated by international organizations, all underpinned by a combination of new and existing legal frameworks in multiple jurisdictions—is slowly emerging, we would do well to learn from the more than two decades of debate in global governance scholarship about the origins, impacts, and shortcomings of related regimes in various other industries, including those not directly connected to digital policy concerns. Because the global AI development landscape is characterized by very high levels of industry control and influence, the question of private power is crucial. How do these firms—and the governance initiatives they create or are a part of—interact with regulatory efforts? How do their efforts to promulgate forms of informal transnational governance interact with their broader political strategies of public influence, lobbying, and interest representation in the global context?

Because the social, economic, and environmental issues AI highlights are not simply results of the technology but entwined with many broader and very different systems, the matter of framing is key. What are the consequences of seeing topics in terms of AI? What other potential forms of governance, both global and otherwise, are being sidelined by global conversations about discrimination, accountability, energy usage, and privacy being seen through the lens of AI, and not otherwise?

As states, companies, interest groups, and other actors have discussions about emerging technologies, what kinds of futures are they setting the scene for, and contributing to willing into being? What implied roles or consequences are there in those futures for those actors, communities, regions, and interests that are not at the table, either because they are being institutionally excluded or because they lack resources, capacity, or connections to be there? If they could set the framing and the tone of the discussion, how different would this conversation look? Bringing it all together, when considering the impact of AI governance initiatives, we would do well to keep in mind the key question raised by critical scholars of international political economy like Susan Strange (Sell 2016, Strange 1998): *Cui bono*? Who really benefits?

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

Aaronson S. 2022. My AI safety lecture for UT Effective Altruism. *Shtetl-Optimized*, Novemb. 28. **https://scottaaronson.blog/?p=6823**

Abbott KW, Snidal D. 2009. The governance triangle: regulatory standards institutions and the shadow of the state. In *The Politics of Global Regulation*, ed. W Mattli, N Woods, pp. 44–88. Princeton, NJ: Princeton Univ. Press

Abelson H, Anderson R, Bellovin SM, Benaloh J, Blaze M, et al. 2021. *Bugs in our pockets: the risks of client-side scanning*. Work. Pap., Columbia Univ., New York. **https://www.cs.columbia.edu/~smb/papers/bugs21.pdf**

Access Now. 2020. *Access Now resigns from the Partnership on AI*. Press Rel., Oct. 13. **https://www.accessnow.org/access-now-resignation-partnership-on-ai/**

Amoore L. 2020. *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*. Durham, NC: Duke Univ. Press

Arun C. 2022. Facebook's faces. *Harvard Law Rev. Forum* 135:236–65

Auld G, Gulbrandsen LH, McDermott CL. 2008. Certification schemes and the impacts on forests and forestry. *Annu. Rev. Environ. Resour.* 33:187–211

Avant DD, Finnemore M, Sell SK. 2010. *Who Governs the Globe?* Cambridge, UK: Cambridge Univ. Press

Balayn A, Gürses S. 2021. *Beyond debiasing: regulating AI and its inequalities*. Rep., Eur. Digit. Rights, Brussels. **https://perma.cc/4UAV-3UFB**

Ballhausen M. 2022. Copyright enforcement. In *Open Source Law, Policy and Practice*, ed. A Brock, pp. 126–40. Oxford, UK: Oxford Univ. Press. 2nd ed.

Bareis J, Katzenbach C. 2022. Talking AI into being: the narratives and imaginaries of national AI strategies and their performative politics. *Sci. Technol. Hum. Values* 47(5):855–81

Bartley T. 2010. Transnational private regulation in practice: the limits of forest and labor standards certification in Indonesia. *Bus. Politics* 12(3):1–34

Belkhir L, Elmeligi A. 2018. Assessing ICT global emissions footprint: trends to 2040 & recommendations. *J. Clean. Prod.* 177:448–63

Bertuzzi L. 2023. US obtains exclusion of NGOs from drafting AI treaty. *Euractiv*, Jan. 17. **https://www.euractiv.com/section/digital/news/us-obtains-exclusion-of-ngos-from-drafting-ai-treaty/**

Bietti E. 2021. From ethics washing to ethics bashing: a moral philosophy view on tech ethics. *J. Soc. Comput.* 2(3):266–83

Binns R, Veale M. 2021. Is that your final decision? Multi-stage profiling, selective effects, and Article 22 of the GDPR. *Int. Data Priv. Law* 11(4):319–32

Birhane A. 2021. Cheap AI. In *Fake AI*, ed. F Kaltheuner, chapter 3. Manchester, UK: Meatspace

Bowker GC, Star SL. 1999. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press

Bradford A. 2020. *The Brussels Effect: How the European Union Rules the World*. New York: Oxford Univ. Press

Braman S. 2011. The framing years: policy fundamentals in the internet design process, 1969–1979. *Inf. Soc.* 27(5):295–310

Bryson JJ. 2020. The artificial intelligence of the ethics of artificial intelligence: an introductory overview for law and regulation. In *The Oxford Handbook of Ethics of AI*, ed. MD Dubber, F Pasquale, S Das, pp. 2–25. Oxford, UK: Oxford Univ. Press

BSR. 2019. *Google celebrity recognition API human rights assessment*. Rep., BSR, New York. **https://www.bsr.org/reports/BSR-Google-CR-API-HRIA-Executive-Summary.pdf**

Caplan R. 2023. Networked platform governance: the construction of the Democratic platform. *Int. J. Commun.* 17:3451–72

Caplan R, Gillespie T. 2020. Tiered governance and demonetization: the shifting terms of labor and compensation in the platform economy. *Soc. Media Soc.* 6(2). **https://doi.org/10.1177/2056305120936636**

Carman A. 2019. What would happen if Apple fully banned Facebook from the App Store? *Verge*, Febr. 1. **https://www.theverge.com/2019/2/1/18205291/apple-facebook-developer-ban-certificate-app-store**

Cashore B, Matus KJ, Norris R. 2012. Pathways to impact: synergies with other approaches. In *Toward Sustainability: The Roles and Limitations of Certification*, ed. Steer. Comm. State-of-Knowledge Assess. Stand. Certif. Washington, DC: Resolve Inc.

Caspers M, Guibault L, McNeice K, Piperidis S, Pouli K, Eskevich M. 2017. *D3.3+ baseline report of policies and barriers of TDM in Europe (extended version)*. Rep., FutureTDM. **https://perma.cc/W6MF-4WLM**

Cath C. 2021. The technology we choose to create: human rights advocacy in the Internet Engineering Task Force. *Telecommun. Policy* 45(6):102144

Cobbe J, Lee MSA, Singh J. 2021. Reviewable automated decision-making: a framework for accountable algorithmic systems. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 598–609. New York: Assoc. Comput. Mach.

Cobbe J, Singh J. 2019. Regulating recommending: motivations, considerations, and principles. *Eur. J. Law Technol.* 10(3). **https://ejlt.org/index.php/ejlt/article/view/686/982**

Cobbe J, Singh J. 2021. Artificial intelligence as a service: legal responsibilities, liabilities, and policy challenges. *Comput. Law Secur. Rev.* 42:105573

Cobbe J, Veale M, Singh J. 2023. Understanding accountability in algorithmic supply chains. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. New York: Assoc. Comput. Mach.

Cohen JE. 2019. *Between Truth and Power: The Legal Constructions of Informational Capitalism*. Oxford, UK: Oxford Univ. Press

Contractor D, McDuff D, Haines JK, Lee J, Hines C, et al. 2022. Behavioral use licensing for responsible AI. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, pp. 778–88. New York: Assoc. Comput. Mach.

Counc. Eur. Union. 2022a. *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (Preparation for COREPER)*. Interinst. File 2021/0106(COD), Counc. Eur. Union, Brussels. **https://perma.cc/564K-RCKR**

Counc. Eur. Union. 2022b. *Council Decision (EU) 2022/2349 of 21 November 2022 authorising the opening of negotiations on behalf of the European Union for a Council of Europe convention on artificial intelligence, human rights, democracy and the rule of law*. Counc. Decis., Counc. Eur. Union, Brussels

Cowls J, Darius P, Santistevan D, Schramm M. 2022. Constitutional metaphors: Facebook's "supreme court" and the legitimation of platform governance. *New Media Soc.* In press

Cowls J, Floridi L. 2019. A unified framework of five principles for AI in society. *Harvard Data Sci. Rev.* 1(1). **https://doi.org/10.1162/99608f92.8cd550d1**

Cowls J, Morley J. 2022. App Store governance: the implications and limitations of duopolistic dominance. In *The 2021 Yearbook of the Digital Ethics Lab*, ed. J Mökander, M Ziosi, pp. 75–92. Cham, Switz.: Springer Int. Publ.

Custers B, Heijne A-S. 2022. The right of access in automated decision-making: the scope of article 15(1)(h) GDPR in theory and practice. *Comput. Law Secur. Rev.* 46:105727

Demetzou K, Zanfir-Fortuna G, Vale SB. 2023. The thin red line: refocusing data protection law on ADM, a global perspective with lessons from case-law. *Comput. Law Secur. Rev.* 49:105806

DeNardis L. 2014. *The Global War for Internet Governance*. New Haven, CT: Yale Univ. Press

Dobbe R, Whittaker M. 2019. AI and climate change: how they're connected, and what we can do about it. *AI Now Institute*, Oct. 17. **https://medium.com/@AINowInstitute/ai-and-climate-change-how-theyre-connected-and-what-we-can-do-about-it-6aa8d0f5b32c**

Douek E. 2019. Facebook's oversight board: move fast with stable infrastructure and humility. *N.C. J. Law Technol.* 21(1):1–78

Dvoskin B. 2022. Expertise and participation in the Facebook oversight board: from reason to will. *Telecommun. Policy* 47(5):102463

Ensmenger N. 2018. The environmental history of computing. *Technol. Cult.* 59(5):S7–33

Erman E, Furendal M. 2022. Artificial intelligence and the political legitimacy of global governance. *Political Stud.* In press

Esty DC. 1994. *Greening the GATT: Trade, Environment, and the Future*. Washington, DC: Inst. Int. Econ.

Finck M. 2018. *Blockchain Regulation and Governance in Europe*. Cambridge, UK: Cambridge Univ. Press. 1st ed.

Funtowicz SO, Ravetz JR. 1993. Science for the post-normal age. *Futures* 25(7):739–55
G20. 2019. *G20 ministerial statement on trade and digital economy*. Statement, G20. **https://www.mofa.go.jp/mofaj/files/000486596.pdf**
Gandy OH. 1993. *The Panoptic Sort: A Political Economy of Personal Information*. Boulder, CO: Westview
Gansky B, McDonald S. 2022. CounterFAccTual: how FAccT undermines its organizing principles. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, pp. 1982–92. New York: Assoc. Comput. Mach.
Goldsmith JL, Wu T. 2006. *Who Controls the Internet? Illusions of a Borderless World*. New York: Oxford Univ. Press
Gorwa R. 2019. The platform governance triangle: conceptualising the informal regulation of online content. *Internet Policy Rev*. 8(2). **https://doi.org/10.14763/2019.2.1407**
Gorwa R, Binns R, Katzenbach C. 2020. Algorithmic content moderation: technical and political challenges in the automation of platform governance. *Big Data Soc*. 7(1). **https://doi.org/10.1177/2053951719897945**
Gray ML, Suri S. 2019. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Boston: Houghton Mifflin Harcourt
Green B. 2019. *The Smart Enough City*. Cambridge, MA: MIT Press
Green B. 2021. The contestation of tech ethics: a sociotechnical approach to technology ethics in practice. *J. Soc. Comput.* 2(3):209–25
Guadamuz A. 2004. Viral contracts or unenforceable documents? Contractual validity of copyleft licenses. *Eur. Intellect. Prop. Rev*. 26(8):331–39
Guadamuz A. 2017. Do androids dream of electric copyright? Comparative analysis of originality in artificial intelligence generated works. *Intellect. Prop. Q.* (2):169–86
Guadamuz A. 2022. DALL·E goes commercial, but what about copyright? *TechnoLlama*, July 25. **https://www.technollama.co.uk/dall%c2%b7e-goes-commercial-but-what-about-copyright**
Harcourt A, Christou G, Simpson S. 2020. Global standard-setting in internet governance. In *Global Standard Setting in Internet Governance*, pp. 1–14. Oxford, UK: Oxford Univ. Press
Jobin A, Ienca M, Vayena E. 2019. The global landscape of AI ethics guidelines. *Nat. Mach. Intell.* 1(9):389–99
Kalyanpur N, Newman AL. 2019. Mobilizing market power: jurisdictional expansion as economic statecraft. *Int. Organ.* 73(1):1–34
Kamara I. 2017. Co-regulation in EU personal data protection: the case of technical standards and the privacy by design standardisation "mandate." *Eur. J. Law Technol.* 8(1). **https://ejlt.org/index.php/ejlt/article/view/545/725**
Keller D. 2021. The future of platform power: making middleware work. *J. Democr.* 32(3):168–72
Kemper J, Kolkman D. 2018. Transparent to whom? No algorithmic accountability without a critical audience. *Inf. Commun. Soc.* 22(14):2081–96
Klonick K. 2019. The Facebook oversight board: creating an independent institution to adjudicate online free expression. *Yale Law J.* 129(8):2418–99
Konisky DM. 2007. Regulatory competition and environmental enforcement: Is there a race to the bottom? *Am. J. Political Sci.* 51(4):853–72
Lambin EF, Thorlakson T. 2018. Sustainability standards: interactions between private actors, civil society, and governments. *Annu. Rev. Environ. Resour.* 43:369–93
Lum K, Isaac W. 2016. To predict and serve? *Significance* 13(5):14–19
Malhotra N, Monin B, Tomz M. 2019. Does private regulation preempt public regulation? *Am. Political Sci. Rev.* 113(1):19–37
Marsden CT. 2011. *Internet Co-Regulation: European Law, Regulatory Governance and Legitimacy in Cyberspace*. Cambridge, UK: Cambridge Univ. Press
Marsden CT, Brown I. 2023. App stores, antitrust and their links to net neutrality: a review of the European policy and academic debate leading to the EU Digital Markets Act. *Internet Policy Rev*. 12(1). **https://doi.org/10.14763/2023.1.1676**
Matus KJM, Veale M. 2022. Certification systems for machine learning: lessons from sustainability. *Regul. Gov.* 16(1):177–96

Munn L. 2022. The uselessness of AI ethics. *AI Ethics*. **https://doi.org/10.1007/s43681-022-00209-w**
Murphy CN. 2015. The last two centuries of global governance. *Glob. Gov.* 21(2):189–96
Nemitz P. 2018. Constitutional democracy and technology in the age of artificial intelligence. *Philos. Trans. R. Soc. A* 376(2133):20180089
Newman A. 2011. Watching the watchers: transgovernmental implementation of data privacy policy in Europe. *J. Comp. Policy Anal.* 13(2):181–94
NIST (Natl. Inst. Stand. Technol.). 2023. *AI Risk Management Framework (AI RMF 1.0)*. Rep., NIST, Washington, DC. **https://doi.org/10.6028/NIST.AI.100-1**
OECD (Organ. Econ. Co-op. Dev.). 2019. *Recommendation on AI*. Recomm. OECD/LEGAL/0449, OECD, Paris. **https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449**
Olney WW. 2013. A race to the bottom? Employment protection and foreign direct investment. *J. Int. Econ.* 91(2):191–203
Pallero J. 2020. *What the Facebook Oversight Board means for human rights, and where we go from here*. Rep., Access Now. **https://apo.org.au/node/307040**
Passi S, Barocas S. 2019. Problem formulation and fairness. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 39–48. New York: Assoc. Comput. Mach.
Perrigo B. 2022. Inside Facebook's African sweatshop. *TIME*, Febr. 17. **https://time.com/6147458/facebook-africa-content-moderation-employee-treatment/**
Perrigo B. 2023. OpenAI used Kenyan workers on less than $2 per hour to make ChatGPT less toxic. *TIME*, Jan. 18. **https://time.com/6247678/openai-chatgpt-kenya-workers/**
Pfotenhauer S, Laurent B, Papageorgiou K, Stilgoe J. 2022. The politics of scaling. *Soc. Stud. Sci.* 52(1):3–34
Png M-T. 2022. At the tensions of South and North: critical roles of Global South stakeholders in AI governance. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, pp. 1434–45. New York: Assoc. Comput. Mach.
Porter G. 1999. Trade competition and pollution standards: "Race to the bottom" or "stuck at the bottom"? *J. Environ. Dev.* 8(2):133–51
Posada J. 2022. Embedded reproduction in platform data work. *Inf. Commun. Soc.* 25(6):816–34
Radauer A, Bader M, Aplin T, Konopka U, Searle N, et al. 2022. *Study on the legal protection of trade secrets in the context of the data economy: final report*. Rep., Eur. Comm., Brussels
Roberts ST. 2019. *Behind the Screen: Content Moderation in the Shadows of Social Media*. New Haven, CT: Yale Univ. Press
Rombach R, Esser P. 2022. *CreativeML Open RAIL-M*. Hugging Face. **https://huggingface.co/spaces/CompVis/stable-diffusion-license**
Sánchez-Monedero J, Dencik L, Edwards L. 2020. What does it mean to "solve" the problem of discrimination in hiring? Social, technical and legal perspectives from the UK on automated hiring systems. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 458–68. New York: Assoc. Comput. Mach.
Schepel H. 2005. *The Constitution of Private Governance: Product Standards in the Regulation of Integrating Markets*. Oxford, UK: Hart
Seaver N. 2017. Algorithms as culture: some tactics for the ethnography of algorithmic systems. *Big Data Soc.* 4(2). **https://doi.org/10.1177/2053951717738104**
Sell SK. 2016. Ahead of her time? Susan Strange and global governance. In *Susan Strange and the Future of Global Political Economy: Power, Control and Transformation*, ed. RD Germain, pp. 21–32. London: Routledge
Shackelford SJ, Proia AA, Martell B, Craig AN. 2015. Toward a global cybersecurity standard of care: exploring the implications of the 2014 NIST Cybersecurity Framework on shaping reasonable national and international cybersecurity practices. *Tex. Int. Law J.* 50(2–3):305–56
Steer. Comm. State-of-Knowledge Assess. Stand. Certif. 2012. *Toward Sustainability: The Roles and Limitations of Certification*. Washington, DC: Resolve Inc.
Stone D. 2019. Transnational policy entrepreneurs and the cultivation of influence: individuals, organizations and their networks. *Globalizations* 16(7):1128–44
Strange S. 1998. Why do international organizations never die? In *Autonomous Policy Making by International Organizations*, ed. B Reinalda, B Verbeek, pp. 213–20. New York: Routledge

ten Oever N. 2021. "This is not how we imagined it": technological affordances, economic drivers, and the Internet architecture imaginary. *New Media Soc.* 23(2):344–62

Tusikov N. 2016. *Chokepoints: Global Private Regulation on the Internet*. Berkeley: Univ. Calif. Press

UNESCO. 2021. *Recommendation on the ethics of artificial intelligence*. Recomm. SHS/BIO/PI/2021/1, UNESCO, Lond. **https://unesdoc.unesco.org/ark:/48223/pf0000381137**

van Doorn N, Badger A. 2020. Platform capitalism's hidden abode: producing data assets in the gig economy. *Antipode* 52(5):1475–95

van Hoboken J, Fathaigh RÓ. 2021. Smartphone platforms as privacy regulators. *Comput. Law Secur. Rev.* 41:105557

van Lente H, Rip A. 2012. Expectations in technological developments: an example of prospective structures to be filled in by agency. In *Getting New Technologies Together*, ed. C Disco, B van der Meulen, pp. 203–30. Berlin: de Gruyter

Vaughan JW, Wallach H. 2021. A human-centered agenda for intelligible machine learning. In *Machines We Trust: Perspectives on Dependable AI*, ed. M Pelillo, T Scantamburlo, pp. 123–38. Cambridge, MA: MIT Press

Veale M. 2020. A critical take on the policy recommendations of the EU High-Level Expert Group on Artificial Intelligence. *Eur. J. Risk Regul.* 11(1):e1

Veale M, Binns R, Edwards L. 2018. Algorithms that remember: model inversion attacks and data protection law. *Philos. Trans. A* 376:20180083

Veale M, Zuiderveen Borgesius F. 2021. Demystifying the Draft EU Artificial Intelligence Act. *Comput. Law Rev. Int.* 22(4):97–112

Voß J-P. 2014. Performative policy studies: realizing "transition management." *Innovation* 27(4):317–43

Wagner B. 2018. Ethics as an escape from regulation: From "ethics-washing" to ethics-shopping? In *Being Profiled: Cogitas Ergo Sum*, ed. E Bayamlıoğlu, I Baraliuc, L Janssens, M Hildebrandt, pp. 84–88. Amsterdam: Amst. Univ. Press

Werner T. 2015. Gaining access by doing good: the effect of sociopolitical reputation on firm participation in public policy making. *Manag. Sci.* 61(8):1989–2011

Whittaker M. 2021. The steep cost of capture. *Interactions* 28(6):50–55

Young M, Katell M, Krafft PM. 2022. Confronting power and corporate capture at the FAccT Conference. *2022 ACM Conference on Fairness, Accountability, and Transparency*, pp. 1375–86. New York: Assoc. Comput. Mach.

Zuboff S. 2015. Big other: surveillance capitalism and the prospects of an information civilization. *J. Inf. Technol.* 30(1):75–89

Annual Review of Law and Social Science

Volume 19, 2023

Contents

Whither Legitimacy? Legal Authority in the Twenty-First Century
Tom R. Tyler 1

Civil Litigants' Evaluations of Their Legal Experiences
Donna Shestowsky 19

Cultivating Equal Minds: Laws and Policies as (De)biasing Social Interventions
Neil A. Lewis Jr. 37

Guantánamo's Legacy
Lisa Hajjar 53

The Ever-Shifting Ground of Pretrial Detention Reform
Jenny E. Carroll 75

Police Go to Court: Police Officers as Witnesses/Defendants
Rachel Moran 93

Centering Race in Studies of Low-Wage Immigrant Labor
Darlène Dubuisson, Patricia Campos-Medina, Shannon Gleeson, and Kati L. Griffith 109

Mandatory Employment Arbitration
Alexander J.S. Colvin and Mark D. Gough 131

Laws of Social Reproduction
Prabha Kotiswaran 145

Firearms Law and Scholarship Beyond Bullets and Bodies
Joseph Blocher, Jacob D. Charles, and Darrell A.H. Miller 165

Beyond Law as a Tool of Public Health: Vaccines in Interdisciplinary Sociolegal and Science Studies
Anna Kirkland 179

Medical Aid in Dying: New Frontiers in Medicine, Law, and Culture
Mara Buchbinder and Cindy Cain 195

How to Study Global Lawmaking: Lessons from Intellectual Property Rights and International Health Emergencies
Tatiana Andia and Nitsan Chorev 215

COVID-19 and the Data Governance Gap
Lisa M. Austin 235

AI and Global Governance: Modalities, Rationales, Tensions
Michael Veale, Kira Matus, and Robert Gorwa 255

Regulating Government AI and the Challenge of Sociotechnical Design
David Freeman Engstrom and Amit Haim 277

How Technology Is (or Is Not) Transforming Law Firms
Ian Rodgers, John Armour, and Mari Sako 299

Bankruptcy Law's Knowns and Unknowns
Jared A. Ellias 319

Refeudalization and Law: From the Rule of Law to Ties of Allegiance
Robert van Krieken 337

Authoritarian Legality and State Capitalism in China
Susan H. Whiting 357

Indexes

Cumulative Index of Contributing Authors, Volumes 1–19 375

Cumulative Index of Article Titles, Volumes 1–19 380

Errata

An online log of corrections to *Annual Review of Law and Social Science* articles may be found at http://www.annualreviews.org/errata/lawsocsci