DATA FOR POLICY CONFERENCE PROCEEDINGS PAPER

A systematic review of regulatory strategies and transparency mandates in AI regulation in Europe, the United States, and Canada

Mona Sloane[1] and Elena Wüllhorst[2]

[1]School of Data Science and Department of Media Studies, University of Virginia, Charlottesville, VA, USA
[2]King's College London, London, UK
Corresponding author: Mona Sloane; Email: mona.sloane@virginia.edu

Abstract

In this paper, we provide a systematic review of existing artificial intelligence (AI) regulations in Europe, the United States, and Canada. We build on the qualitative analysis of 129 AI regulations (enacted and not enacted) to identify patterns in regulatory strategies and in AI transparency requirements. Based on the analysis of this sample, we suggest that there are three main regulatory strategies for AI: AI-focused overhauls of existing regulation, the introduction of novel AI regulation, and the omnibus approach. We argue that although these types emerge as distinct strategies, their boundaries are porous as the AI regulation landscape is rapidly evolving. We find that across our sample, AI transparency is effectively treated as a central mechanism for meaningful mitigation of potential AI harms. We therefore focus on AI transparency mandates in our analysis and identify six AI transparency patterns: human in the loop, assessments, audits, disclosures, inventories, and red teaming. We contend that this qualitative analysis of AI regulations and AI transparency patterns provides a much needed bridge between the policy discourse on AI, which is all too often bound up in very detailed legal discussions and applied sociotechnical research on AI fairness, accountability, and transparency.

Policy Significance Statement

This study of existing regulatory strategies for artificial intelligence (AI) as well as patterns of AI transparency mandates within the current landscape of AI regulation in Europe, the United States, and Canada will inform the global AI regulation discourse as well as sociotechnical research on compliance with AI transparency mandates.

1. Introduction

AI systems increasingly pervade social life, commerce, and the public sector. AI systems include autocorrect and text suggestion features in word processing systems, automated risk assessments and algorithmic trading, or predictive analytics used for resource allocation planning in government agencies. In 2023, this development was accelerated by the release of text- and image-based generative AI systems such as OpenAI's ChatGPT, Google's Bard, or Anthropic's Claude. AI, yet, again, is making headlines.

Over the past years, evidence has been mounting that AI systems can cause harms and amplify already existing systems of discrimination and oppression (SafiyaUmojaNoble, 2018; Eubanks, 2017; Buolamwini & Gebru, 2018; Bolukbasi et al., 2016). These include limiting access to resources and opportunity, healthcare, or loans as well as oversurveillance and wrongful accusations of criminal behavior, among others (Johnson, 2022; Ledford, 2019).

At the same time, however, AI is seen as a conduit for innovation and prosperity, leading to governments' large scale investments in AI across a range of domains, such as research, innovation, training, and national security (S, 2022a; Hunter et al., 2018; European Commission, 2017) and underlining AI's growing significance in the current phase of geopolitics. This phase is often portrayed as featuring a "global AI race" (Kokas, 2023; Annoni et al., 2018) in which the most powerful geopolitical powers (chiefly the United States, China, and the European Union) compete for global AI leadership. This includes large-scale data collection and trading, model development, as well as semiconductor production and chips innovation (Miller, 2022).

AI is not new, but has—in its basic form—been around for many years. AI systems can broadly be defined as "a class of computer programs designed to solve problems requiring inferential reasoning, decision-making based on incomplete or uncertain information, classification, optimization, and perception" (Bathaee, 2018). Broad distinctions can be made between rule-based and learning-based AI, with the former being given explicit instructions (such as what symptoms lead to a diagnosis or conditional statements based on the rules of chess) to follow and the latter being fed labeled or unlabeled data and independently detecting patterns in the data (Stoyanovich & Khan, 2021). Generative AI are learning-based systems that ingest vast amounts of raw data to learn to generate statistically probable outputs when prompted by encoding simplified representations of the ingested data to produce new data that is similar but not identical to the training data (Martineau, 2023a).

As civil society, policymakers and governments, as well as industry players, are increasingly made aware of the potential harms of AI systems, including harms that can flow from generative AI, they are under pressure to devise regulations and governance measures that mitigate potential AI harm: AI *policy* has entered the "AI race" (Espinoza et al., 2023). Most recently, this race has been accelerated by the re-invigoration of "existential risk" narratives in the AI discourse that posit that generative AI is proof of AI's potential to overthrow humanity (Roose, 2023). These narratives have been propelled by prominent and powerful AI actors, such as Geoffrey Hinton who famously declared in mid-2023 that the potential threat of AI to the world was a more urgent issue than climate change (Coulter, 2023). Although subsequently framed as "AI safety" and distinctly different from the calls of academics and activists for the mitigation of real-world AI harms that are already occurring (versus being an artifact of the future), the release of generative AI into global markets and its accompanying "existential risk" discourse elevated AI regulation to a public concern.

1.1. The global race for AI regulation

Unsurprisingly, AI and technology policy scholarship is now concerned with structuring and analyzing evolving AI policy discourse. While this is challenging due to its rapidly evolving nature—by September 2023, 190 new state-based AI regulations had been introduced in the United States alone (BSA | The Software Alliance, 2023), and in late October 2023, President Biden had signed the sweeping "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence" (US President, 2023)—there are general meta trends in AI *regulation* that are discernible. These trends map onto the cultural and political differences that characterize leading geopolitical powers and, therefore, also mirror geopolitical divisions. For example, the United States tend to follow a market-driven regulatory model which focuses on incentives to innovate and centers markets over government regulation (A, 2023). This is a contrast to the Chinese state-driven regulatory model that sets out to deploy AI to strengthen government control (Kokas, 2023) and the European model which emphasizes government oversight and rules (A, 2023).

The geopolitical tensions around AI also influence how AI regulation is framed and rationalized. For example, the United States legislature regulates AI research and deployment as a national security concern

(Fischer et al., 2021). This includes state-wide bans of foreign AI and data collection products, such as the Chinese social media services TikTok and WeChat (Governor Glenn Youngkin, 2022), or the call for such bans on a US federal level (Kang & Maheshwari, 2023; Protecting Americans from Foreign Adversary Controlled Applications Act, 2024).

While the race for AI regulation has been interpreted as fruitful for identifying best policies and regulatory tools (Smuha, 2021), there is also an absence in global harmonization of AI governance, an issue that the UK AI Safety Summit held in November 2023 sought to address (Smuha, 2023). However, an emphasis on these global perspectives of AI regulation conceals the efforts to regulate AI at different levels of government, creating a gap in the literature on subnational AI policy (Liebig et al., 2022).

1.2. AI regulation and transparency

Another gap in scholarship emerges on the "other side of the coin" of AI regulation: compliance. The body of work on strategies for compliance specifically with AI regulation is nascent. For the most part, there is agreement on the vagueness of many AI regulations (Schuett, 2023), although enforcement potentials always depend on the regulatory set-up. At the same time, policy and legal scholars have turned to examining potential adaptions and uses of existing governance mechanisms to AI applications. Examples include the need to adapt the definition of protected groups, as codified in antidiscrimination legislation, to better harmonize with the opaque ways algorithms sort people into groups (Wachter, 2022), or the evaluation of existing governance measures, such as impact assessments and audits that are common in other domains, for AI systems (Stahl et al., 2023; Nahmias & Perel, 2020; D et al., 2020).

While enhancing AI "safety" and ensuring harm mitigation are *goals* cutting across most AI regulations on global, national, and subnational levels, there is a central mechanism that is assumed to enable the *achievement* of those goals: the "opening up" of AI's technical and legal "black box" (Pasquale, 2016; Bathaee, 2018) by way of mandating different forms of AI *transparency* (Larsson & Heintz, 2020). In other words, AI transparency is often considered cornerstone of safe and equitable AI systems (Ingrams & Klievink, 2022; Novelli et al., 2023; Mittelstadt et al., 2016), even though it is messy in practice (F et al., 2019; Bell et al., 2023) and by no means a panacea (Ananny & Crawford, 2018; Cucciniello et al., 2017) since universally perfect AI transparency is unachievable (Sloane et al., 2023).

AI transparency, however, is as much of a legal and regulatory concern as it is a technical compliance concern and, therefore, is a sociotechnical concern. The kinds of transparency measures that are mandated by regulators in AI regulations ultimately must be translated into action by AI industry and enforcement agencies. This challenge is not new to computer, data, and information science disciplines where a concern with applied methods for opening up AI's "black box" and developing "fair, accountable and transparent" machine learning techniques has become a rapidly growing field over the past 5 years. (This is, for example, evidenced in the rapid growth of the Fairness, Accountability and Transparency (FAccT) Conference of the Association for Computing Machinery (ACM): accepted papers have grown from 63 papers in 2018 to 503 in 2022 [Stanford University Human-Centered Artificial Intelligence, 2023].) This field, however, is not necessarily informed by a close reading of the ongoing AI regulation discourse. There is a disconnect between applied research on fair, accountable, and transparent AI and the ways in which the AI policy discourse frames and treats the same issues. In other words, AI researchers working on social impact issues from a sociotechnical standpoint (especially outside of corporate research labs) are not always in tune with the regulatory discourse, missing out on the potential to provide alternative pathways to AI compliance that are not directed by corporate structures and interests (S, 2022b; Young et al., 2022). This disconnect means that vast potentials for synergistic work between research and policy spaces on AI transparency are foreclosed.

2. Study

To address the disconnect described above, an analysis of the ways in which AI regulation is strategically approached and how AI transparency is framed *within* AI regulation is required. This is particularly true for

countries in which both AI industry and regulatory activity are strong. Prompted by that, we survey the landscape of proposed and enacted AI regulations in Europe, the United States, and Canada—geographies that house both growing AI industries and demonstrate high AI regulatory activity—and provide a qualitative analysis of regulatory strategies and patterns in AI transparency. Our work is driven by the following research questions:

- What are regulatory *strategies* in existing AI regulations in Europe, the United States and Canada?
- What are the *patterns of AI transparency measures* in existing AI regulation in Europe, the United States and Canada?

Rather than assessing the impact or effectiveness of AI regulation, our work is focused on analyzing the regulations themselves. The goal is to provide insight into how regulations *express* strategies for AI harms mitigation and *articulate* transparency requirements and what patterns can be discerned in both regards. This foregrounds the ways in which AI harms mitigation and AI transparency techniques become *meaningful* within the transnational regulatory discourse.

2.1. Data collection

Our data collection focused on AI regulations (in this paper, we use a broad frame for AI regulations; we include any legally binding frameworks that govern AI in society [which includes bills, laws, ordinances, and more]) in the European Union and European Countries, in the United States, and in Canada. To be included in the sample, AI regulations in our chosen geographies had to be either enacted or not enacted. The latter included proposed regulations, inactive regulations, that is, regulations that were introduced but not ratified, stalled, as well as not yet reintroduced regulations. To be included, regulations also had to explicitly address either the functioning or the operation of AI systems, or both. We base this distinction on the following definition of transparency: Existing works in governance scholarship tend to frame transparency as either transparency of *artefacts*, such as budgets, or transparency of *processes*, such as decision-making (Cucciniello et al., 2017; Heald, 2006; Meijer et al., 2018; Ingrams & Klievink, 2022). We build on this two-pronged approach (transparency of both artifacts and processes; including both an artefact- and process- (or use)-based approach is particularly salient in the context of regulatory debates about generative AI which can be considered "general purpose technology" (Caliskan & Lum, 2024)) to define AI transparency as mandates for accessibility and availability of information on the functioning of AI systems and their operation. Based on this definition, we exclude regulations exclusively focused on establishing committees or councils as they are neither focused on the functioning of AI systems, nor their operation. Additionally, AI-adjacent umbrella regulations such as data privacy regulations and fiscal laws mentioning AI, but not explicitly addressing AI, were excluded.

Across the three geographies, we followed the AI localism approach (Verhulst et al., 2021) to distinguish AI regulations based on their local, state, national, and transnational reach. Consequently, agency- and government-internal AI regulations and processes are excluded since they do not articulate AI regulations on a local, state, national, or transnational level. Based on this, government resolutions are also excluded, except when they establish an AI ban or moratorium, that is, they explicitly target AI. Additionally, two exceptions were made: the 2022 US Blueprint for an AI Bill of Rights which was added due to its signaling effect for AI regulation across the United States; and similarly US Executive Orders on AI which do impact how US agencies engage with AI.

For our data collection, we relied on existing AI policy repositories that are freely accessible. Specifically, the OECD.AI policy observatory (OCED, 2023); the AI Watch website by the European Commission's Joint Research Centre (JRC) (European Commission's Joint Research Center, 2023); the collection of legislation related to AI by the National Conference of State Legislatures in the United States (NCSL, 2023) (which was last updated in September 2023); the US Chamber of Commerce AI Legislation Tracker (U.S. Chamber of Commerce, 2022); and the map of US locations that banned facial recognition by Fight for the Future (Fight for the Future, 2023). Additionally, we used internet search engines, specifically Google, with the following

keywords and combinations thereof: "government artificial intelligence regulation," "federal AI regulation," *plus specific years such as* "2023 AI regulation," "2023 local AI regulation," and "2023 European AI regulation."

The data collection took place in two phases. Phase 1 consisted of an initial collection taking place from June to August 2022. The second data collection phase took place from September to October 2023. Both searches followed the same methodology outlined above. All regulations meeting the criteria outlined above were collected in the bibliography software Zotero.

Overall, we collected a total of 138 AI regulations in Europe, the United States, and Canada. After removing nine duplicates (duplicates are stalled or inactive regulations that were reintroduced verbatim in the same chamber or in a different chamber, or were introduced verbatim in a different state; duplicates also include same-text regulations that are pending in more than one chamber), the final dataset was comprised of 129 AI regulations. All AI regulations in the sample were proposed or enacted in the time frame from July 2016 to October 2023. The AI regulatory *formats* featured in the sample were directives (directives are regulations which define a goal that must be legislated by the respective lower level of government, for example, directives of the European Commission must be implemented by the member states), amendments, acts, bills (which are US-specific), municipal codes, ordinances, administrative codes, and moratoria.

The AI localism distribution across the sample is as follows: 22% of all regulations are on the local level, 53% are on the state level, 23% are on the national level, and 2% are on the transnational level (see Figure 1).

The dataset is heavily skewed toward US regulations with 90% of all regulations in the dataset being US regulations (a total of 116 regulations), only 8% European regulations (a total of 11 regulations), and 2% (a total of 2 regulations) Canadian regulations. The United States also demonstrates the biggest diversity in terms of AI localism. Figure 2 shows the regulatory activity in the United States. National regulations apply to the whole country. Green-colored states indicate that at least one regulation has been enacted on the state level. Yellow-colored states indicate that at least one regulation has been proposed on the state level. Finally, green pins indicate that at least one regulation has been enacted on the local level. Certain clusters of local regulations can be observed both in Massachusetts, where there are eight local regulations in the state alone, and California where there are seven local regulations in the Bay Area. The regulations enacted on the local level often share content in that they use similar formulations and specify requirements sometimes verbatim. For example, the regulations on surveillance technology in Oakland, Berkeley, Davis and Santa Clara County, California center on the same requirements (surveillance use
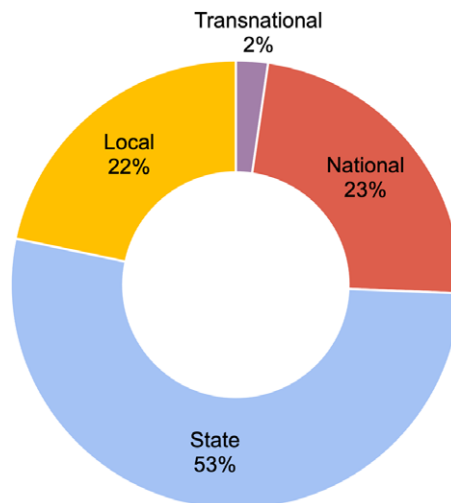

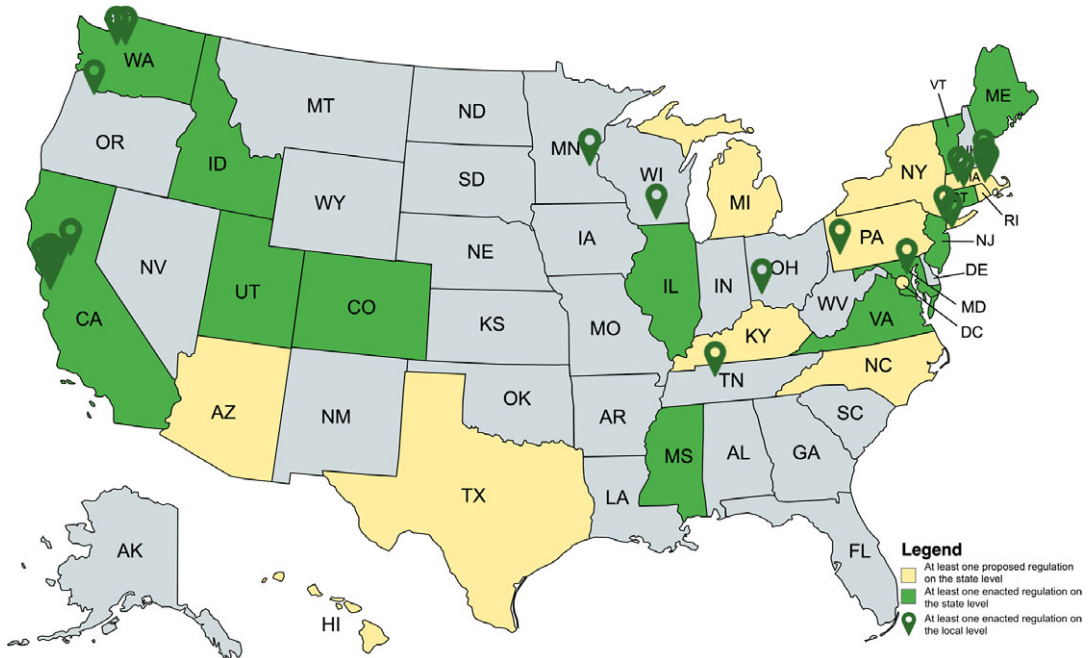***Figure 1.*** *AI localism distribution across the dataset.*

Figure 2. *Map of regulatory activity in the United States on the state and local level.*

policy and annual surveillance report) (Regulations on City's Acquisition and Use of Surveillance Technology, 2018; Municipal Code, 2018; Surveillance Technology Ordinance, 2018; Surveillance – Technology and Community – Safety Ordinance, 2016).

2.2. Data analysis

Our data analysis followed a qualitative coding approach designed to identify patterns in semantic meaning. We used both descriptive codes (i.e., codes that summarize a topic with one word, often verbatim derived from the data, such as "amendment") and evaluative codes (i.e., codes that summarize the interpretation of the data, such as "transnational" or "human in the loop") (Saldaña, 2021). Our data analysis and coding practice was driven by the following prompts:

- What is the type of regulation is the AI regulation?
- What are the regulatory strategies or concrete AI governance measures proposed in the regulation?
- What are the AI transparency requirements, if any, articulated in the regulation?

Based on these guiding questions, we read each regulation and wrote a summary of the regulation and of the transparency requirements contained within it. Additionally, we qualitatively coded each regulation. The qualitative coding process was comprised of three steps. First, existing categories or labels contained in the regulations were descriptively coded *as patterns*. For example, if the regulation was called "amendment to XXX," then it was coded as "amendment" (descriptive coding). Second, the scope of the regulation was analyzed, that is, the concern of the regulation with the functioning of AI, and/or specific sectors and their use of AI technologies and applications (evaluative coding; using codes such as "facial recognition technology"). Third, the explicit governance measures described in the regulation were analyzed (evaluative coding; using codes such as "ban"). To identify patterns in AI transparency requirements, the same strategy was deployed, using a combination of descriptive and evaluative codes to analyze how AI transparency was framed and articulated in the regulations. During the coding process, patterns began to emerge immediately for both the regulatory strategies and the transparency mandates.

Type	Description	Patterns within the regulations of this type	Example Regulation
Overhauls	Overhauls involve updating existing legal frameworks pertaining to AI-focused regulations to account for AI-specific risks and impacts.	There are three main types of AI-focused regulations that often are overhauled: • Non-discrimination legislation • Consumer protection legislation • Road traffic law	An example of an overhauled, AI-specific regulation is the proposed supplement to the 1945 New Jersey Law Against Discrimination. The proposed supplement seeks to recognize the discrimination occurring in automated decision systems by establishing protected groups. Members of the protected groups would be safeguarded against discrimination by automated decision systems, in areas such as banking, insurance, and healthcare. *This supplement has been proposed but has not been enacted yet.*
Novel AI Regulations	Novel AI Regulations are newly developed regulations focused exclusively on AI.	Novel AI Regulations often focus on: • A specific type of AI technology, such as facial recognition technology • A specific application of AI, such as the use of AI in employment decisions • AI literacy and skill enhancement	Novel AI Regulations that focus on, for example, AI literacy and skills, mandate regulations of skill assessments, as well as enhancements of specific segments of the workforce. An example of a Novel AI Regulation is the proposed 2022 New Jersey Act requiring the Commissioner of Labor and Workforce Development to conduct a study and issue a report about the impact of AI on the growth of New Jersey's economy. *This regulation has been proposed but has not been enacted yet.*
Omnibus Regulations	Omnibus Regulations are large, not sector specific, and may contain more than one substantive matter.	Omnibus Regulations set out to regulate AI itself, not the industry or application of AI. Omnibus Regulations typically take a long time to be developed and passed.	An example of an Omnibus Regulation is the EU AI Act, proposed in 2018, that deploys a risk-based approach by tailoring AI technology requirements to the specific risks associated with the use of said AI technology. *A provisional agreement was reached between the European Parliament and Council in December 2023.*

Figure 3. *The three dominant regulatory strategies with examples.*

We discussed these emerging patterns in weekly meetings and began to consolidate codes based on their growing overlap in meaning. For example, we consolidated all codes that described governance measures exclusively focused on types of AI technology, or areas of application into "Novel AI Regulations," one of the regulatory strategies we identified as a major pattern. Similarly, we combined all codes that described some form of human intervention as AI transparency into "Human in the Loop."

2.3. Limitations

As with any research, this study has a number of limitations. First, its focus on AI regulations limits what can be said about the political and social processes surrounding AI law-making. For example, excluding an analysis of the discourse of AI regulation, as well as ethnographic work on the processes of creating AI regulation excludes the generation of knowledge about whose voices and concerns are heard and addressed in AI regulation, and whose are ignored. In other words, this research does not address issues of inequality in AI regulation. Furthermore, albeit working toward creating synergies between AI policy, social science research, and applied AI work (including compliance), this study does not include research on how technologists deal with AI regulation, or even contribute to it through work on fair, accountable and transparent AI, nor does it include research on compliance practice or the "careers" of AI regulations over time.

Second, our dataset shows a number of biases that limit our ability to make generalizable statements based on our analysis. Our dataset is characterized by a Western-centric view, excluding regulatory developments on AI in other geographies and cultures, including those with very active AI and software industries, such as China or India, and those historically excluded from global technology discourses, such as African countries. Additionally, the heavy skew toward US regulations is a limitation. Furthermore, the European sample is heavily biased due to the language limitations of the research team, who only were able to access and process regulations written in English or German. The European regulations consist of AI regulations at the level of the European Union, as well as British, Austrian, German, and Dutch AI regulations. (It must be noted at this point that from very early on AI regulation was considered a matter of the European Union, rather than national legislation. The European Union began working on the topic of AI regulation as early as 2017, a development effectively halting AI regulations on the national
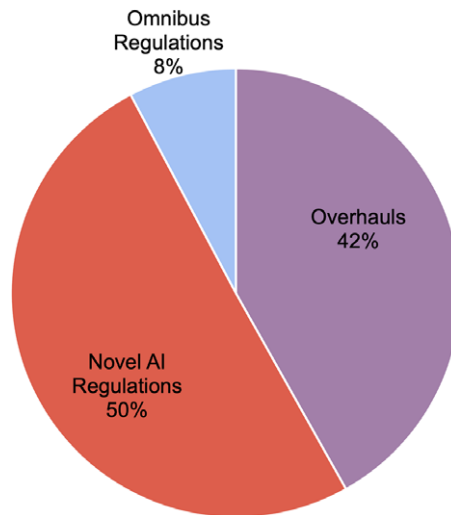
Figure 4. *Distribution of regulatory strategies across the dataset.*

level of EU member states.) It must also be noted that 55% of all regulations in the dataset are not enacted yet or are inactive. In other words, this study is an examination of *aspirations* and strategic directions of the AI regulatory discourse, not an assessment of the general effectiveness of AI regulations.

3. Findings and discussion

Our data analysis identified three types of dominant regulatory strategies that cut across AI regulation in Europe, the United States, and Canada, as well as six AI transparency patterns.

3.1. Regulatory strategies

The three types of dominant regulatory strategies for AI are overhauls, novel AI regulations, and the omnibus approach.

In our sample, 42% of the regulations were overhauls (a total of 54 regulations), 50% were novel AI regulations (a total of 64 regulations), and 8% were omnibus regulations (a total of 11 regulations). This can be seen in Figure 4 (in percentages).

It must be noted that even though omnibus regulations present a minority in the sample, omnibus regulations generally affect more people than novel AI regulations or overhauls of existing regulations with respect to AI. This is because omnibus regulations tend to occur at the national or transnational level. The EU has taken an omnibus and transnational approach to AI regulation through the EU AI Act. In contrast, the United States shows a significant amount of regulatory activity on the state level. But fewer people are affected by US state-level activity. This means that the omnibus approach pursued in Europe affects more people on average than the more fragmented approach in the United States. (In the Appendix, we provide population data per regulation.)

3.1.1. Overhauls

The first regulatory strategy that emerges at the intersection of AI and lawmaking is overhauling existing legal frameworks with a focus on AI. The driving question, here, is how existing regulatory frameworks, policies, and laws can be leveraged to account for AI-specific risks and impacts and mitigate those.

Notable examples of AI-focused overhauls are specific adjustments or additions that are made to US anti-discrimination legislation. For example, the proposed supplement to the 1945 New Jersey Law Against Discrimination (this supplement was introduced twice, once in 2020 (S 1943)(Gill, 2020) and

once in 2022 (S 1402)(Gill, 2022)), which seeks to modify existing anti-discrimination legislation to account for discrimination that can occur in the use of automated decision systems or AI. Under this supplement, members of protected groups would be safeguarded against discrimination by automated decision systems, particularly in areas like banking, insurance, and healthcare. The amendment was proposed but not enacted in 2020 (S 1943) (Gill, 2020) and was recently reintroduced unchanged.

Another area in which AI-focused overhauls occur is consumer protection. Some overhauls strengthen user and consumer rights online, in particular with respect to large online platforms and their use of content moderation practices and targeted advertisement technologies, some of which can be AI-driven. For example, the California Business and Professions Code (SB 1001) (Bots, 2018) was amended in 2018 to mandate the disclosure of bots in commercial transactions.

In Europe, there have been AI-specific overhauls, too. For example, the Dutch 2016 Amendment of the Road Traffic Act (Amendment of the Road Traffic Act, 2016) specifies that road tests of autonomous vehicles are exempt from the Road Traffic Act (Wegenverkeerswet, 1994) if the experiments are monitored and evaluated by the Dutch vehicle authority. Similarly, the German 2017 Eight Act amending the Road Traffic Act (Gesetz zum automatisierten Fahren, 2017) amends the "Straßenverkehrsgesetz" (Road Traffic Act) (Road Traffic Act (Straßenverkehrsgesetz) (Section 7–20), 2021) and details rules and requirements for both manufacturers and drivers of vehicles with a highly or fully automated driving function.

3.1.2. Novel AI regulations

The second dominant regulatory strategy for AI is the development of entirely novel regulations that are exclusively focused on AI. These come at different scales and levels but are always new laws created within a given legal regime (such as a code of ordinances, which is a city's compilation of all its laws and some of its regulations). Within this type of strategy, a pattern emerges: novel AI regulations tend to focus on the artefact of AI, that is, a specific *type of AI technology*, such as facial recognition technology, on a *specific application or use of AI*, such as the use of AI in employment or healthcare, or they set out to enhance AI literacy and skills.

An example of novel AI regulations focused on a specific AI technology is the growing body of regulations focusing on facial recognition technology. In the United States, there are many local facial recognition technology regulations. Most are concerned with the use of facial recognition technology by local government agencies or departments. Local governments either enforce a moratorium (Moratorium on Facial Recognition Technology, 2020; Moratorium on Facial Recognition Technology in Schools, 2023) or a complete ban of facial recognition technology, such as the City of Northampton in Massachusetts (Prohibition on the Use of Face Recognition Systems by Municipal Agencies, Officers, and Employees, 2019), which enacted a prohibition of the use of face surveillance systems by municipal agencies, officers, and employees in 2019. More often the ban comes with a partial exemption for law enforcement like in Alameda, California (Prohibiting the use of Face Recognition Technology, 2019), or a full exemption for law enforcement, such as in Maine (HP 1174) (ILAP, 2021). Municipal governments also often specify an approval process for facial recognition or, more generally, surveillance technology.

Another focus in AI regulations directed at a specific AI technology are autonomous vehicles. Respective AI regulations have been enacted in the UK (Automated and Electric Vehicles Act, 2018) and Austria (Automatisierte Fahrsysteme, 2016), typically specifying conditions for the testing of vehicles with automated driving functions. Often there are specific requirements for military vehicles and again different requirements pertaining to different levels of automation within the vehicle.

An example of novel AI regulations targeting the use of AI in a specific domain are regulations concerned with the use of automated systems, or AI, in hiring and employment. For example, the Illinois Artificial Intelligence Video Interview Act (820 ILCS 42) (Artificial Intelligence Video Interview Act, 2020) requires that applicants must give consent to employers seeking to use AI analysis on recorded video interviews. Similarly, New York City's Local Law 144 (The New York City Council, 2021) requires that applicants are notified of any AI use that may occur in the application process. It also mandates annual bias audits for automated employment decision tools. Similarly, a proposed regulation in Pennsylvania targets the use of AI in health insurance, mandating the disclosure of AI use in the review process

(HB 1663) (Rep Venkat, 2023). Furthermore, AI regulation often targets government procurement and use of AI specifically, such as a proposed Vermont act relating to state development, use, and procurement of automated decision systems (H 263) (Brian Cina, 2021).

Novel AI regulations designed to enhance AI literacy and skills are, for example, regulations that mandate skill assessments and the enhancement of specific segments of the workforce. An example of this is the proposed 2022 New Jersey Act requiring the Commissioner of Labor and Workforce Development to conduct a study and issue a report on the impact of artificial intelligence as it pertains to the growth of the state's economy (A 168) (Carter, 2022).

3.1.3. Omnibus regulations

Regardless of the domain, an omnibus regulation is generally large in scope, not constrained to specific sectors, and may contain more than one substantive matter. An example of an omnibus approach to data protection is the European General Data Protection Regulation (GDPR) (S, 2019). Omnibus approaches to regulation are rare. They typically take a long time to develop and get passed. For example, the GDPR took 8 years, from 2011 to 2019, to be fully developed and come into effect (European Data Protection Supervisor, 2023). The groundwork for the EU AI Act was laid in 2018 with the launch of a high-level expert group working on AI. In the United States, various regulations following an omnibus approach have been introduced over the years, chiefly the proposed Algorithmic Accountability Act of 2019 (S 1108) (Wyden and Booker, 2019) and the proposed Algorithmic Accountability Act of 2022 (S 3572) (Senator Wyden, 2022). As of Spring 2024, both the European Union and the United States have omnibus approaches for putting guardrails around AI. In the European Union, it is the EU AI Act (the EU AI Act was introduced in April 2021. The European Parliament and Council reached a provisional agreement on the Artificial Intelligence Act on December 9th 2023) (Artificial Intelligence Act (EU AI Act), 2024), and in the United States, it is the Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (EO 14110; the US President manages the operations of the executive branch of government through Executive Orders; Executives Orders are not the law but are prescriptive for the operations of government agencies (Executive Orders, 2024)) (US President, 2023). While the EU AI Act is a regulation, the Executive Order only directs the design and use of AI in the US executive branch.

These frameworks set out to regulate AI *itself*, not the industry or application. To achieve that, the EU AI Act (Artificial Intelligence Act (EU AI Act), 2024) deploys a risk-based approach. The Canadian Directive on Automated Decision-Making (Directive on Automated Decision-Making, 2023) deploys a risk-based approach, too. However, like the US Executive Order, it only addresses *government* use of AI. The EU's risk-based approach involves tailoring requirements to the specific risks associated with the use of each AI technology. The EU AI Act introduces a tiered risk approach that categorizes the risk of AI use as either unacceptable, high, limited, or minimal and imposes respective "ex-ante" and "ex-post" testing and mitigation approaches, chiefly premarket deployment impact assessments and postmarket deployment audits (Mökander et al., 2022).

The US Executive Order (EO 14110) (US President, 2023) is not based on one single approach to regulating AI, such as a risk-based approach, but assembles various concerns and approaches under one umbrella. It is centered on the notion of AI safety, security, and trustworthiness and articulates how these are to be achieved across eight areas of concern (The White House, 2023): AI safety and security standards; privacy; equity and civil rights; consumer protection; the labor market; innovation and competition; international collaboration; and government use of AI. Implementation and enforcement of the Executive Order is also decentralized and falls onto different federal agencies.

3.1.4. Discussion of regulatory strategies

This typology of regulatory strategy cuts across the AI localism levels since any of these strategies are deployed at the transnational, national, state, and local level. Figure 5 shows what types of regulation were proposed between 2016 and 2023 at each AI localism level, indicating trends in regulatory strategies. We can see that on the state level, there is a large increase in proposals to overhaul existing regulations with respect to AI. Additionally, an overall increase in proposed AI regulations AI can be observed. Similarly,
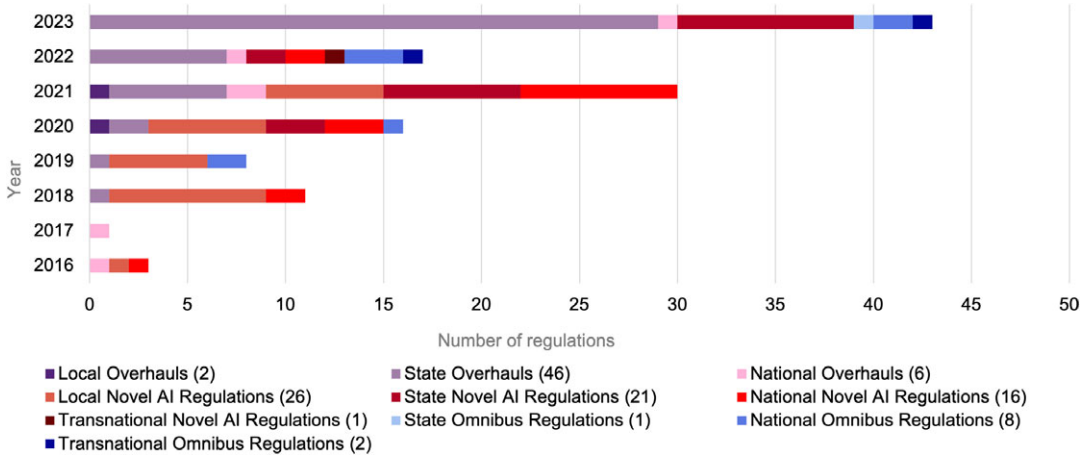
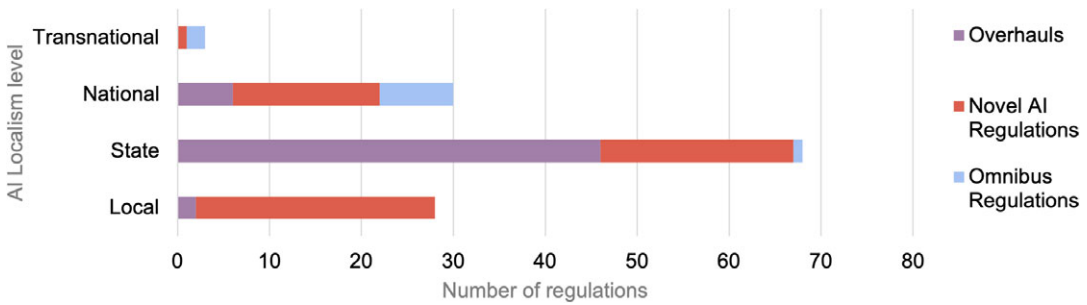**Figure 5.** *Regulatory strategies in absolutes by AI localism level per year.*

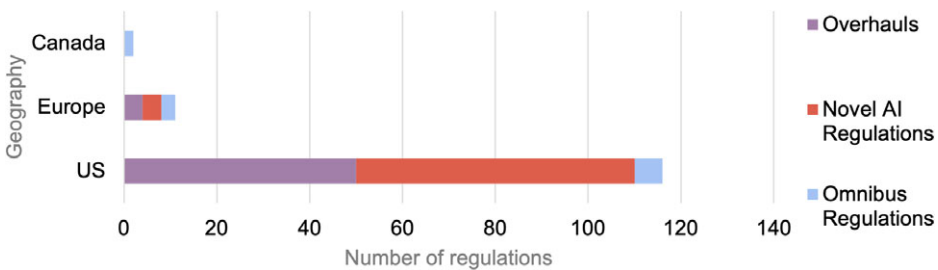**Figure 6.** *Regulatory strategies in absolutes by AI localism level.*

Figure 7. *Regulatory strategies in absolutes by geography.*

Figure 6 shows regulatory strategies per AI Localism level. It can be observed that on the state and local level, regulations are generally novel AI regulations or overhauls. Omnibus regulations are typically proposed at the transnational or national level.

Figure 7 shows the prevalence of the regulatory strategies across geographies. As indicated earlier, European and Canadian regulations focus on omnibus regulations, whereas in the United States, predominately overhauls and novel AI regulations are proposed. Figure 7 also shows the skew in our sample toward US regulations, as mentioned in the limitation section.

It is important to note that the boundaries between the categories of regulatory strategies we introduced are extremely porous. AI regulation strategies continually evolve and increasingly overlap. For example, they may have the characteristics of an omnibus approach but focus on a particular sector, for example by virtue of their policy reach. Examples of this are both the Canadian Directive on Automated Decision-

Making (Directive on Automated Decision-Making, 2023) and the US Executive Order on Safe, Secure and Trustworthy AI (EO 14110) (US President, 2023) which both focus on AI broadly, but only in terms of government use of AI.

Similarly, AI regulations may focus on AI technology alongside the application of AI. For example, there are AI regulations that are AI technology-specific, as well as application-specific, but that are an overhaul, such as the 2023 proposed overhaul of the Arizona revised statues relating to the conduct of elections which targets the use of AI in machines used for elections, for the processing of ballots, or the electronic vote adjudication process (S 1565; the act was vetoed by the Governor of the State of Arizona Katie Hobbs in April 2023 (Hobbs, 2023)) (Carroll, 2023). Similarly, a new AI regulation may build on the omnibus approach but be dedicated to AI upskilling and innovation, like the 2019 US Executive Order on Maintaining American Leadership in Artificial Intelligence (EO 13859) (US President, 2019).

The patterns in AI regulation approaches and their dynamics per geography and AI Localism level map onto similar studies presented by other scholars. For example, Marcucci et. al (Marcucci et al., 2023) find a fragmented and emergent ecosystem of global data governance and, among other aspects, identify dominant patterns in terms of principles, processes, and practice. Their recommendation to work toward a comprehensive and cohesive global data governance framework, however, targets policy makers rather than sociotechnical researchers. Similarly, Dotan (Dotan, 2024) surveys the evolving landscape of US AI regulation, covering much longer time periods, and finds a fragmented field in which regulatory approaches increasingly converge.

3.2. *AI transparency patterns*

The AI transparency mandates identified in the sample fall into six categories. They are: human in the loop, assessments, audits, disclosures, inventories, and red teaming. Figure 8 shows the distribution of individual transparency mandates in the sample: 8% are human in the loop requirements, 35% are assessments, 27% are disclosure requirements, 16% are audits, 13% are inventories, and 1% are red teaming.

In total, there are 157 individual transparency mandates assigned across all regulations in the sample. This includes regulations with combinations of transparency mandates. Overall, 48 out of the 129 regulations analyzed do not assign any transparency mandate.

3.2.1. *Human in the loop*

Many regulations treat human intervention as a pathway to mitigating potential AI harms. The requirements for human in the loop typically mandate that either a human must supervise high-risk decisions
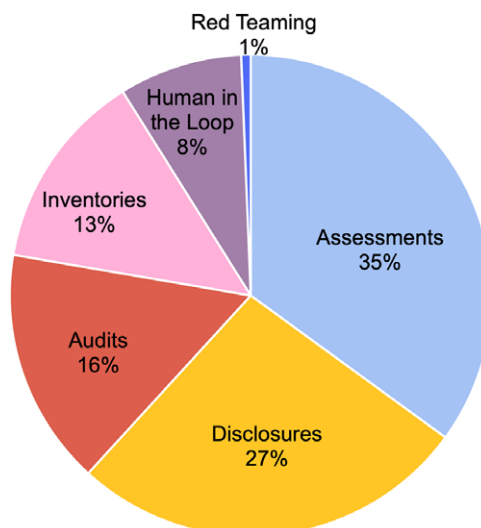

Figure 8. *Distribution of transparency mandates across the dataset.*

and/or be able to intervene, or an affected individual can demand human intervention. This usually requires at least a partial opening up of AI's "black box" because the system has to be designed in a way that a human actually *can* intervene.

Example: An example of this is an IT Deployment Law, enacted in April 2022 in Schleswig-Holstein, Germany (IT Deployment Law (IT-Einsatz-Gesetz), 2022). This state regulation governs the use of data-driven information technology by public entities. Of particular interest is the law's Human in the Loop provision which requires that in cases of technical failure, at the request of the affected person, or if the system fails to produce a result, human intervention must replace the technology. The law stipulates that there must be a clearly defined process enabling humans to deactivate the technology and assume control of the task.

3.2.2. Assessments

Assessments are techniques designed to evaluate the systemic risks and impacts of AI. These are typically articulated in terms of civil rights and liberties, privacy, and discrimination. Assessments can be mandated to be completed both premarket deployment (ex-ante) or postmarket deployment (ex-post). Some AI regulations do not only mandate AI assessments, but also propose risk mitigation measures (Digital Services Oversight and Safety Act, 2022). Assessments feature prominently in most AI regulations.

Example: An example of mandated assessments is the mandatory use of the Algorithmic Impact Assessment (AIA) tool (Treasury Board of Canada, 2023) in the Canadian Directive on Automated Decision-Making (Directive on Automated Decision-Making, 2023). This tool is designed as a questionnaire to help assess the impact level of an automated decision-making system. It consists of questions focused on risk alongside questions on mitigation. The assessment scores are derived from various factors such as the system's architecture, its algorithm, the nature of decisions it makes, the impact of these decisions, and the data it uses.

3.2.3. Audits

Audits are a prominent feature in many AI regulations. They can also be considered an ex-post assessment (which is the case in the EU AI Act). Audits generally compare nominal values or information to actual values or information (Sloane & Moss, 2023), but they are often not clearly defined in AI regulation. Where they are specified, however, they are deployed to identify violations of policies or laws. Here, the comparisons of how an AI should behave (nominal value) versus how it actually behaves (actual value; this includes audits of potential bias in AI training data), or of how it should be used versus how it is actually used, serve as a technique for identifying violations of policies or laws, such as anti-discrimination law, consumer protection regulation, or specific use policies.

Example: New York City's Local Law 144 (The New York City Council, 2021) mandates an annual bias audit. Even though the regulation does not define the term 'bias audit', it is widely interpreted in the context of US anti-discrimination legislation in employment. That means that employment and hiring AI is checked for potential adverse impacts on individuals and communities based on protected categories such as race, religion, sex, national origin, age, disability, or genetic information. Local US regulations concerning surveillance technology often mandate both a surveillance use policy and an annual report following the deployment of such technology. Audits, included as part of the annual report, serve as a mechanism to ensure adherence to the surveillance use policy. A notable instance of this can be seen in Oakland, California's regulations on the city's acquisition and use of surveillance technology [85], which were enacted in May 2018 and subsequently amended in January 2021.

3.2.4. Disclosures.
Disclosure requirements typically prescribe that AI users must be notified that an AI system is at work. In general, AI regulations mandating disclosure state that the disclosure format must be a clear, conspicuous notice in plain language. Some AI regulations require disclosure for bots, advertisements, automated decision systems, and content moderation. Disclosure requirements can be combined with other requirements, such as human in the loop.

Examples: An example is an amendment of the California Business and Professions Code relating to bots, enacted in 2018 (SB 1001) (Bots, 2018). Bots are defined as online accounts whose actions and posts are not produced by a person. When bots are used to communicate or interact in commercial transactions

Transparency Mandate	Description	Example Regulation	Geography	AI Localism Level	Description of Transparency Mandate
Human in the Loop	Human-in-the-Loop requirements state that high-risk decision-making that utilizes AI must be supervised by a human, or users of an AI system must be given the option to request human intervention or assessment of said AI system.	IT Deployment Law (IT-Einsatz-Gesetz) *enacted 2022*	Schleswig Holstein, Germany	State	The regulation concerns the use of data-driven information technology by public entities. It follows a risk-based approach. Every decision involving a data-driven information system must be attributable to a human. If a system makes a decision without being initiated by a human, a human must take over when: • there is a technical failure • the person affected by the decision requests human consideration ("KI Rüge") • the automated system does not produce a result
Assessments	Assessments evaluate technology's systemic risks and impacts on civil rights and liberties, privacy, and discrimination. Assessments can be mandated ex-ante and ex-post.	The Directive on Automated Decision-Making *enacted 2019, latest modification: 2023*	Canada	National	The Directive on Automated Decision-Making is a regulatory framework for automated decision-making systems that requires algorithmic impact assessments to be completed before the production of an automated decision-making system begins. These assessments must be reviewed regularly. Moreover, the Treasury Board of Canada created a publicly available algorithmic impact assessment tool. The directive this tool is based on requires users to conduct a peer review of the automated decision-making system dependent on its impact level as classified by the algorithmic impact assessment tool.
Audits	Audits can be considered an ex-post assessment and are often not defined in regulations. When defined, audits are put in place to identify violations of policies or laws.	A local Law to amend the administrative code of the city of New York, in relation to automated employment decision tools (Local Law 144) *enacted 2021*	New York City, New York, United States	Local	Local Law 144 enforces audit and notice requirements for automated employment decision tools. An automated employment decision tool is a "computational process, derived from (..) artificial intelligence, that issues simplified output, including a score, classification, or recommendation, that is used to substantially assist or replace discretionary decision making for making employment decisions". The Law prohibits the use of automated employment decision tools unless an audit checking for bias was conducted within a year prior to its use. A bias audit is defined as an "impartial evaluation by an independent auditor" and involves testing the automated employment decision tool for disparate impacts. The results of the bias audit must be made publicly available on the employer's website.
Disclosures	Disclosures are a notification that an artificial intelligence system is in use. Disclosures are required to be clear, conspicuous, and in plain language.	An Act to add Chapter 6 to Part 3 of Division 7 of the Business and Professions Code, relating to bots *enacted 2018*	California, United States	State	The Act mandates the disclosure of the use of bots in commercial transactions and when bots have been used to influence election voting. A bot is defined as an "online account where all (..) actions or posts of that account are not the result of a person". The disclosure must be clear, conspicuous, and in a comprehensible format.
Inventories	Inventories are a comprehensive overview of the use cases of a specific technology. They can contain information, such as a description of the function and use of a technology, as well as the person or company liable for the technology.	An Act to increase privacy and security by prohibiting the use of facial surveillance by certain government employees and officials (HP 1174) *enacted 2021*	Maine, United States	State	The Act prohibits the use of face surveillance systems, except during the investigation of a crime. All requests to access face surveillance data by public officials must be recorded and stored publicly in de-identified logs. These logs must include the: • Date and name of public employee making the request • Databases searched • Statutory offense under investigation • Race and sex of the person under investigation
Red Teaming	Red Teaming is a strategy used to test the vulnerabilities and weaknesses of a system where a designated red team within an organization initiate attacks on a system in a controlled environment.	Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (EO 14110) *US President Biden 2023*	United States	National	Executive Order 14110 was signed by US President Joe Biden in October 2023, and Red Teaming is a prominent part of the Executive Order. The EO requires that the director of NIST must establish guidelines for developers of AI who conduct Red Teaming tests. Specifically, companies developing dual-use foundation models, which are self-supervised models containing "at least tens of billions of parameters," to conduct Red Teaming tests must report the results of the Red Teaming tests to the Federal Government. The report must include mitigation steps that were taken after the tests were completed. Furthermore, when AI is used by the Federal Government, the Director of the Office of Management and Budget must provide guidance on the external tests of AI systems, including Red Teaming tests of generative AI systems.

***Figure 9.** The six types of AI transparency mandates with examples.*

or in an election then this must be disclosed in a clear, conspicuous, and reasonably designed format. Similarly, a proposed 2023 AI regulation in Illinois (HB 1002) (Illinois Hospital Act, 2023) combines the disclosure requirement with the option of human consideration (i.e., human in the loop): when diagnostic AI is used, it must be disclosed and patients can request a human assessment instead.

3.2.5. Inventories

Inventories are designed to provide an overview of the uses of AI by specific organizations or agencies. Their format varies. An inventory can provide information on the training data (a popular format for providing information on training data is datasheets (Gebru et al., 2021); datasheets are, for example, mandated in the EU AI Act; here, training data sets must be described (including provenance, scope and main characteristics, labeling, and cleaning), alongside training methodologies and techniques (Artificial Intelligence Act (EU AI Act), 2024)) and the functioning and use of AI systems, as well as information on liability. Inventories are not always public, nor are they necessarily comprehensive.

Example: An example is the California Act concerning automated decision systems (AB 302) (Ward, 2023) which was approved by the Governor in October 2023. This regulation requires the Department of Technology to conduct an inventory of all AI systems that are classified as "high-risk" (high-risk automated decision systems in this regulation are defined as automated decision systems that assist or replace human discretionary decisions that have a legal or similarly significant effect (Ward, 2023)) and have been proposed or are already developed, procured, or used by any state agency prior to September 1, 2024. The inventory must include information on how the AI supports or makes decisions, its benefits and potential alternatives, evidence of research on assessing its efficacy and benefits, the data and personal information used by the AI, and risk mitigation measures. (This example shows that there are intersections between governance measures. The inventory includes a risk assessment.)

Another regulation requiring an inventory of use cases of surveillance technology is Maine Act HP 1174 to increase privacy and security by prohibiting the use of facial surveillance by certain government employees and officials (ILAP, 2021). The regulation was enacted in July 2021. It prohibits government departments, employees, and officials from using face surveillance systems and their data except for specific purposes. One exception is the investigation of a crime. Requests to search facial surveillance systems must be logged by the Bureau of Motor Vehicles and the State Police. The logs are public records and must contain demographic information, for example the race and sex of the person under investigation.

3.2.6. Red teaming

Red teaming is a relatively novel addition to the canon of AI transparency patterns. It originates in military simulations in the Cold War where adversarial teams were assigned the color red (Ji, 2023). Red teaming has since been employed as a tactic in cybersecurity. Typically, red teaming is performed by internal engineers of a company (although not always, see the example below) with the goal of testing specific outcomes of certain types of interactions with a system, such as adversarial attacks (such as obtaining access to personal credit card details). The goal of red teaming is to identify (otherwise hidden) flaws in the system so that they can then be addressed (Friedler et al., 2023). Red teaming is currently considered to be a valid addition to already existing AI (risk) assessment processes (Friedler et al., 2023).

Example: An example is the 2023 Executive Order on AI (EO 14110) (US President, 2023) since it includes AI red teaming as a key measure. In August 2023, the White House also cohosted a large-scale generative AI red team competition at the hacker conference DEF CON (Ji, 2023). Outside of these instances, red teaming has not formally been proposed in any of the regulations surveyed for this study.

3.2.7. Discussion of transparency mandates

Like the patterns in AI regulation strategies, the AI transparency patterns often intersect and combine. For instance, regulations may mandate risk assessments alongside bias audits. To illustrate the popularity of individual transparency mandates and their combinations, Figure 10 shows transparency mandates in the regulations in absolute numbers. The diagonal entries showcase the number of regulations assigning a single transparency mandate: 55 regulations in the sample require some form of assessment. The off-diagonal entries showcase the pairings of transparency mandates. There are, for example, 24 regulations that require both assessment and disclosure. This heat map is symmetric. Following a temperature scale, red color indicates high and dark blue low frequency.

We want to note that the six transparency mandates we identified in our dataset as dominant should not be taken as comprehensive, nor as best practice. The field of fair, accountable and transparent AI is
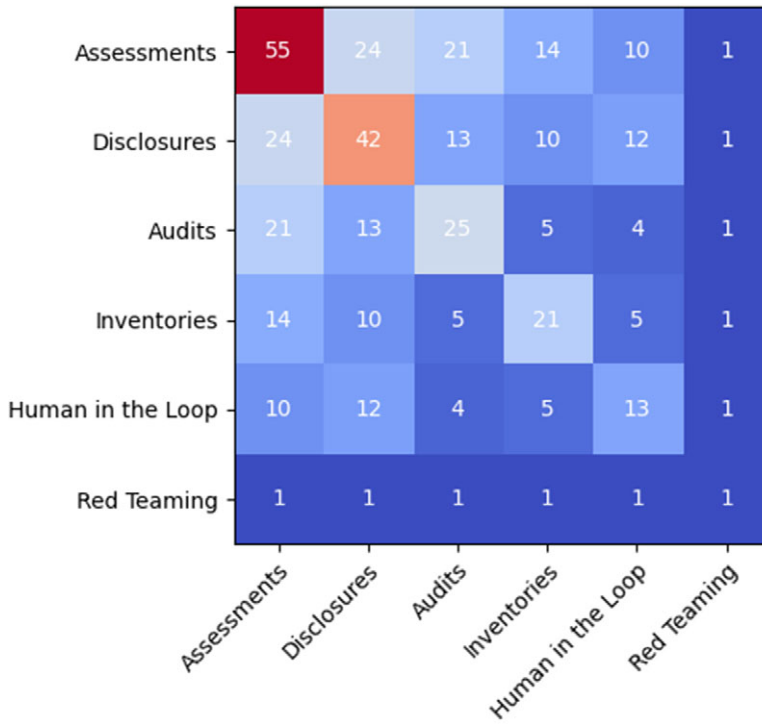
Figure 10. *Heat map of transparency mandates.*

constantly evolving. The introduction of generative AI has brought about new challenges to AI accountability and transparency, but also new solutions that cut across transparency and model improvement. For example, retrieval-augmented generation (RAG) is a new technique that directs large language models (LLMs) to retrieve facts from an external knowledge base (for example, a specific text corpus) to ensure higher accuracy and up-to-date information, as well as provide transparency about the output generation (Martineau, 2023b). Users can check an LLM's claims vis-à-vis the base corpus. RAG cannot exclusively be classified as transparency technique; it also is an accuracy technique. Current regulations do not account for these new developments. Moving forward, regulators may want to consider the growing overlaps between compliance and innovation as they refine AI regulations and the transparency mandates contained within them. This is particularly true for many of the more specific state regulations that are not enacted yet.

We also want to note that the regulatory strategies and transparency mandates we outline do not automatically mitigate the structural issues that persist in the realm of AI design and society at large. These include, but are not limited to, pervasive AI harms, homogeneous groups entrusted with AI design and governance and resulting power imbalances, the extractive nature of data collection for AI systems, or AI's ecological impact.

Against that backdrop, we want to caution against using the list of AI transparency mandates in a check box manner. Rather, we envision a future of applied socio-technical research where researchers can build on these patterns to push forward with applied research on any of these mandates in order to avoid corporate capture of compliance. Empowering the research community to push forward with this effort in an interdisciplinary way (rather than letting corporate legal teams decide on what good compliance looks like) keeps the door open for asking questions around the structural issues that AI might be involved in and that AI regulation set out to mitigate in the first place.

4. Conclusion and Future Work

In this paper, we provided a bottom-up analysis of AI regulation strategies and AI transparency patterns in Europe, the United States, and Canada. The goal of this study was to begin building a bridge between the

AI policy world and applied socio-technical research on fairness, accountability and transparency in AI. The latter is a research space in which methods for compliance with new AI regulations will emerge. We envision future socio-technical research to explicitly build on the patterns of AI transparency mandates we identify and develop meaningful compliance techniques that are public interest- rather than corporate interest-driven, and that are research-based. These may include more holistic audit approaches that include assumptions, stakeholder-driven impact assessments and red teaming approaches, profession-specific disclosures, comprehensive and mandatory inventories, and meaningful human in the loop processes. We hope that this study can be a resource to further this work and make the AI policy discourse, which often is complex and legalistic, more accessible.

In addition to the analytical work contained in the paper, we hope that our work will prompt not only pragmatic, but also critical research on the global landscape of AI regulation. In particular, we hope that future work will shed light on the AI regulations that were not included in this study, particularly in geographies that tend to be under-represented in AI research, and in research more generally. We also hope that future research turns to ethnographic explorations of the AI law-making process in different cultural contexts in order to generate a better understanding of how the typologies we identify in this paper come about, and what their politics are. Lastly, we hope that future work will examine how participatory policy-making and participatory compliance processes can be established.

References

Bradford A (ed.) (2023) *Digital Empires*. New York, NY: Oxford University Press.

Amendment of the Road Traffic Act (2016) *Amendment of the Road Traffic Act 1994 in connection with the enabling of experiments with automated systems in vehicles.* Available at: https://www.internetconsultatie.nl/experimenteerwet_zelfrijdendeauto/details (accessed 7 June 2022).

Ananny M and Crawford K (2018) Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society 20*(3), 973–989.

Annoni A *et al.* (2018) *Artificial Intelligence: A European Perspective.* Available at: https://publications.jrc.ec.europa.eu/repository/handle/JRC113826 (accessed 21 November 2023).

Artificial Intelligence Act (EU AI Act) (2024). Available at: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html (accessed 26 November 2023).

Artificial Intelligence Video Interview Act (2020). Available at: https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=4015&ChapterID=68 (accessed 3 June 2022).

Automated and Electric Vehicles Act (2018) Publisher: King's Printer of Acts of Parliament. Available at: https://www.legislation.gov.uk/ukpga/2018/18/contents/enacted (accessed 24 September 2023).

Automatisierte Fahrsysteme (2016) Automatisiertes Fahren Verordnung (Austrian Automated Driving Regulation). Available at: https://www.ris.bka.gv.at/Dokumente/BgblAuth/BGBLA_2016_II_402/BGBLA_2016_II_402.pdfsig (accessed 7 June 2022).

Bathaee Y (2018) The artificial intelligence black box and the failure of intent and causation. *Harvard Journal of Law & Technology 31*(2) 889–938.

Bell A, **Nov O and Stoyanovich J** (2023) Think about the stakeholders first! Toward an algorithmic transparency playbook for regulatory compliance. *Data & Policy 5*, e12.

Bolukbasi T *et al.* (2016) Man Is to Computer Programmer as Woman Is to Homemaker? Debiasing Word Embeddings. Available at: http://arxiv.org/abs/1607.06520 (accessed 18 November 2023).

Bots 2018 An act to add Chapter 6 (commencing with Section 17940) to Part 3 of Division 7 of the Business and Professions Code. Available at: https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180SB1001 (accessed 20 June 2022).

Brian Cina (2021) An act relating to State development, use, and procurement of automated decision systems – pending. Available at: https://legislature.vermont.gov/bill/status/2022/H.263 (accessed 3 June 2022).

BSA | The Software Alliance 2023 *BSA Analysis: State AI Legislation Surges by 440% in 2023.* Available at: https://www.bsa.org/news-events/news/bsa-analysis-state-ai-legislationsurges-by-440-in-2023 (accessed 18 November 2023).

Buolamwini J and Gebru T (2018) Gender shades: intersectional accuracy disparities in commercial gender classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*. Conference on Fairness, Accountability and Transparency, January 212018. PMLR, pp. 77–91.

Caliskan A and Lum K (2024) *Effective AI Regulation Requires Understanding General-Purpose AI*. Brookings. Available at: https://www.brookings.edu/articles/effective-ai-regulation-requires-understanding-general-purpose-ai/ (accessed 24 March 2024).

Carroll (2023) An act amending sections 16-442, 16-552 and 16-621, Arizona Revised Statutes, relating to conduct of Elections (S 1565). Available at: https://custom.statenet.com/public/resources.cgi?id=ID:bill:AZ2023000S1565&ciq=ncsl&client_md=009f86572673187364f38c98b0425a99&mode=current_text (accessed 24 September 2023).

Carter 2022 An act requiring the commissioner of labor and workforce development to conduct a study and issue a report on the impact of artificial intelligence on the growth of the State's economy. Available at: https://custom.statenet.com/public/resources.cgi?id=ID:bill:NJ2022000A168&ciq=ncsl&client_md=e0bcd2dc37ab5c2e1f3e6ce6a4e15f82&mode=current_text (accessed 26 September 2023).

Coulter M (2023) AI pioneer says its threat to world may be 'more urgent' than climate change. *Reuters.* Available at: https://www.reuters.com/technology/ai-pioneer-says-itsthreat-world-may-be-more-urgent-than-climate-change-2023-05-05/ (accessed 18 November 2023).

Cucciniello M, **Porumbescu GA and Grimmelikhuijsen S** (2017) 25 Years of transparency research: Evidence and future directions. *Public Administration Review 77*(1), 32–44.

Raji Inioluwa D *et al.* (2020) Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. FAT* '20: Conference on Fairness, Accountability, and Transparency, January 27, 2020. Barcelona, Spain: Association for Computing Machinery, pp. 33–44.

Digital Services Oversight and Safety Act (2022) Lori [D-MA-3 Rep. Trahan. Archive Location: 2022-02-18. Available at: https://www.congress.gov/bill/117thcongress/house-bill/6796/text (accessed 9 September 2023).

Directive on Automated Decision-Making (2023). Available at: https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592 (accessed 26 November 2023).

Dotan R (2024) US regulation of artificial intelligence. In: Alexander K, Christoph L and Raphael M (eds), *The Handbook on Applied AI Ethics*. Edward Elgar Publishing: Cheltenham, United Kingdom 153–178.

Espinoza J, **Criddle C, and Liu Q** (2023) The global race to set the rules for AI. *Financial Times.* Available at: https://www.ft.com/content/59b9ef36-771f-4f91-89d1-ef89f4a2ec4e (accessed 18 November 2023).

Eubanks V (2017) *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, 1st Edn. New York, NY: St. Martin's Press, p. 260.

European Commission (2017) *Harnessing the Economic Benefits of Artificial Intelligence*. Digital Transformation Monitor. Available at: https://ati.ec.europa.eu/reports/technology-watch/harnessing-economic-benefits-artificial-intelligence (accessed 27 November 2023).

European Commission's Joint Research Center (2023) *AI Watch*. European Commission. Available at: https://ai-watch.ec.europa.eu/index_en (accessed 21 November 2023).

European Data Protection Supervisor (2023) *The History of the General Data Protection Regulation.* Available at: https://edps.europa.eu/data-protection/data-protection/legislation/history-general-dataprotection-regulation_en (accessed 26 November 2023).

Executive Orders (2024) Federal Register. Available at: https://www.federalregister.gov/presidential-documents/executive-orders (accessed 27 March 2024).

Heike F *et al.* (2019) "Transparency you can trust: transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society 6*(1) 1–14.

Fight for the Future (2023) *Ban Facial Recognition Map.* Available at: https://www.banfacialrecognition.com/map/ (accessed 21 November 2023).

Fischer SC *et al.* (2021) *AI Policy Levers: A Review of the U.S. Government's Tools to Shape AI Research, Development, and Deployment*. Centre for the Governance of AI, Future of Humanity Institute, University of Oxford: Oxford, United Kingdom.

Friedler S *et al.* (2023) *AI Red-Teaming Is Not a One-Stop Solution to AI* Harms: Data & Society.

Gebru T *et al.* (2021) Datasheets for datasets. *Communications of the ACM 64*(12), 86–92.

Gesetz zum automatisierten Fahren (2017). Änderung des Straßenverkehrsgesetzes. Available at: https://bmdv.bund/de/Share dDocs/EN/Documents/DG/eight-act-amending-the-road-traffic-act.pdf?__blob=publicationFile (accessed 11 June 2024).

Gill (2020) An act concerning discrimination and automated decision systems and supplementing P.L.1945, c.169 (C.10:5-1 et seq.). Available at: https://www.njleg.state.nj.us/bill-search/2020/S1943 (accessed 3 June 2022).

Gill (2022) An act concerning discrimination and automated decision systems and supplementing P.L.1945, c.169 (C.10:5-1 et seq.). Available at: https://custom.statenet.com/public/resources.cgi?id=ID:bill:NJ2022000S1402&ciq=ncsl&client_md=32b714614d9a84996599f9aadae3257c&mode=current_text (accessed 24 September 2023).

Governor Glenn Youngkin (2022) Executive order banning the use of certain applications and websites on state government technology. Available at: https://www.governor.virginia.gov/newsroom/news-releases/2022/december/name-948259-en.html (accessed 18 November 2023).

Heald D (2006) Varieties of transparency. In: Hood C and Heald D (eds), *Transparency: The Key to Better Governance?*. Oxford University Press Inc: New York, United States of America, 24–43.

Hobbs G (2023) Governor Veto SB 1565. Available at: https://www.azleg.gov/govlettr/56leg/1r/sb1565.pdf (accessed 24 March 2024).

Hunter AP *et al.* (2018) Artificial intelligence and national security: the importance of the AI ecosystem. Available at: https://www.csis.org/analysis/artificial-intelligence-andnational-security-importance-ai-ecosystem (accessed 27 November 2023).

ILAP (2021) An act to increase privacy and security by prohibiting the use of facial surveillance by certain government employees and officials. Available at: http://www.mainelegislature.org/legis/bills/getPDF.asp?paper=HP1174&item=2&snum=130 (accessed 17 June 2022).

Illinois Hospital Act (2023) ***Amends the University of Illinois Hospital Act (regarding diagnostic algorithms)***. Available at: https://legiscan.com/IL/text/HB1002/id/2637549 (accessed 9 October 2023).

Ingrams A and Klievink B (2022) Transparency's role in AI governance. In: Bullock JB *et al.* (eds), *The Oxford Handbook of AI Governance*. Oxford University Press: New York, United States of America.

IT Deployment Law (IT-Einsatz-Gesetz) (2022). Available at: https://www.gesetze-rechtsprechung.sh.juris.de/jportal/portal/t/s4u/page/bsshoprod.psml;jsessionid=2909846397D32668116D3C1C0E43BF1F.jp29?pid=Dokumentanzeige&showdoccase=1&js_peid=Trefferliste&fromdoctodoc=yes&doc.id=jlr-ITEGSHRahmen&doc.part=X&doc.price=0.0#focuspoint (accessed 5 July 2022).

Ji J (2023) What does AI red-teaming actually mean? Center for Security and Emerging Technology. Available at: https://cset.georgetown.edu/article/what-does-ai-red-teaming-actually-mean/ (accessed 23 November 2023).

Johnson K (2022) How wrongful arrests based on AI derailed 3 men's lives. *Wired.* Available at: https://www.wired.com/story/wrongful-arrests-aiderailed-3-mens-lives/ (accessed 18 November 2023).

Kang C and Maheshwari S (2023) Lawmakers renew calls to ban TikTok after accusations of anti-Israel content. *The New York Times.* Available at: https://www.nytimes.com/2023/11/08/business/tiktok-accusations-anti-israel-content.html (accessed 18 November 2023).

Kokas A (2023) *Trafficking Data: How China Is Winning the Battle for Digital Sovereignty.* New York, NY: Oxford University Press, p. 335.

Larsson S and Heintz F (2020) Transparency in artificial intelligence. *Internet Policy Review 9*(2) 1–16.

Ledford H (2019) Millions of black people affected by racial bias in health-care algorithms. *Nature 574*(7780), 608–609.

Liebig L *et al.* (2022) Subnational AI policy: shaping AI in a multi-level governance system. *AI & Society 39*, 1477–90.

Marcucci S *et al.* (2023) Informing the global data future: benchmarking data governance frameworks. *Data & Policy 5*, e30.

Martineau K (2023a) *What Is Generative AI?* IBM Research Blog. Available at: https://research.ibm.com/blog/what-is-generative-AI?p1=Search&p4=43700078077908955&p5=p&gclsrc=aw.ds (accessed 18 August 2023).

Martineau K (2023b) What is retrieval-augmented generation? IBM Research Blog. Available at: https://research.ibm.com/blog/retrieval-augmented-generation-RAG (accessed 27 March 2024).

Meijer A, **Paul 't H and Worthy B** (2018) Assessing government transparency: an interpretive framework. *Administration & Society 50*(4), 501–526.

Miller C (2022) *Chipwar: The Fight for the World's Most Critical Technology.* First Scribner Hardcover Edition. New York, NY: Scribner, p. 431.

Mittelstadt BD *et al.* (2016) The ethics of algorithms: mapping the debate. *Big Data & Society 3*(2) 1–21.

Mökander J *et al.* (2022) Conformity assessments and post-market monitoring: a guide to the role of auditing in the proposed European ai regulation. *Minds and Machines 32*(2), pp. 241–268.

Moratorium on Facial Recognition Technology (2020). Available at: https://legislature.vermont.gov/Documents/2022/WorkGroups/House%20Judiciary/Bills/H.195/Drafts,%20Amendments,%20and%20Other%20Legal%20Documents/H.195~Michele%20Childs~2020%20Acts%20and%20Resolves%20No.166~2-12-2021.pdf (accessed 17 June 2022).

Moratorium on Facial Recognition Technology in Schools (2023). Available at: https://www.nysed.gov/sites/default/files/programs/data-privacy-security/biometricdetermination-9-27-23.pdf (accessed 26 November 2023).

Municipal Code (2018) Acquisition and Use of Surveillance Technology, Chapter 2.99 – Berkeley, California, US. Available at: https://berkeley.municipal.codes/BMC/2.99 (accessed 17 June 2022).

Nahmias Y and Perel M (2020) *The Oversight of Content Moderation by AI: Impact Assessments and Their Limitations*. Rochester, NY. Available at: https://papers.ssrn.com/abstract=3565025 (accessed 18 September 2023).

NCSL (2023) *Artificial Intelligence 2023 Legislation*. National Center of State Legislatures (NCSL). Available at: https://www.ncsl.org/technology-and-communication/artificial-intelligence-2023-legislation (accessed 21 November 2023).

Novelli C, **Taddeo M and Floridi L** (2023) Accountability in artificial intelligence: What it is and how it works. *AI & Society 39*, 1871–1882.

OCED 2023 *Emerging AI-Related Regulation*. OCED.AI policy observatory. Available at: https://oecd.ai/en/dashboards/policy-instruments/Emerging_technology_regulation (accessed 21 November 2023).

Pasquale F (2016) *The Black Box Society: The Secret Algorithms That Control Money and Information*. First Harvard University Press paperback edition. Cambridge: Harvard University Press, p. 311.

Prohibiting the use of Face Recognition Technology (2019). Available at: http://docs.alamedaca.gov/WebLink/DocView.aspx?id=645835&dbid=0&repo=CityofAlameda&cr=1 (accessed 17 June 2022).

Prohibition on the Use of Face Recognition Systems by Municipal Agencies, Officers, and Employees (2019). Available at: https://www.northamptonma.gov/AgendaCenter/ViewFile/Item/13774?fileID=130290 (accessed 17 June 2022).

Protecting Americans from Foreign Adversary Controlled Applications Act (2024). Available at: https://www.congress.gov/bill/118th-congress/house-bill/7521/text?s=1&r=1.

Regulations on City's Acquisition and Use of Surveillance Technology (2018) City Code Chapter 9.64-Oakland, California, US. Available at: https://library.municode.com/ca/oakland/codes/code_of_ordinances?nodeId=TIT9PUPEMOWE_CH9.64REACUSSUTE (accessed 17 June 2022).

Rep Venkat (2023) An act providing for disclosure by health insurers of the use of artificial intelligence-based algorithms in the utilization review process. Available at: https://www.legis.state.pa.us/cfdocs/billinfo/billinfo.cfm?syear=2023&body=H&type=B&bn=1663 (accessed 10 October 2023).

Road Traffic Act (Straßenverkehrsgesetz) (Section 7–20) (2021). Available at: https://www.gesetzeim-internet.de/englisch_stvg/index.html (accessed 26 November 2023).

Roose K (2023) AI poses "risk of extinction," industry leaders warn. *The New York Times.* Available at: https://www.nytimes.com/2023/05/30/technology/aithreat-warning.html (accessed 18 November 2023).

Mona S (2022a) *Threading Innovation, Regulation, and the Mitigation of Ai Harm: Examining Ethics in National AI Strategies*. Rochester, NY. Available at: https://papers.ssrn.com/abstract=4143375 (accessed 31 October 2023).

Mona S (2022b) To make AI fair, here's what we must learn to do. *Nature 605*(7908), 9.

Paul MS (2019) Global data privacy: the EU way. *New York University Law Review 94*(4) 771–818.

SafiyaUmojaNoble (2018) *Algorithms of Oppression: How Search Engines Reinforce racism*. New York, NY: New York University Press.

Saldaña J. (2021) *The Coding Manual for Qualitative Researchers*, 4th Edn. Thousand Oaks, CA: SAGE, p. 414.

Schuett J (2023) Defining the scope of AI regulations. *Law, Innovation and Technology 15*(1), 60–82.

Senator Wyden (2022) Algorithmic Accountability Act of 2022. Available at: https://www.congress.gov/bill/117th-congress/senate-bill/3572/text (accessed 9 September 2023).

Sloane M and Moss E (2023) *Assessing the Assessment: Comparing Algorithmic Impact Assessments and AI Audits*. Rochester, NY. Available at: https://papers.ssrn.com/abstract=4486259 (accessed 26 November 2023).

Sloane M *et al.* (2023) Introducing contextual transparency for automated decision systems. *Nature Machine Intelligence 5*(3), 187–195.

Smuha NA (2021) From a "race to AI" to a "race to AI regulation": regulatory competition for artificial intelligence. *Law, Innovation and Technology 13*(1), 57–84.

Smuha NA (2023) *Biden, Bletchley, and the Emerging International Law of AI*. Verfassungsblog. Available at: https://verfassungsblog.de/biden-bletchley-and-the-emerginginternational-law-of-ai/ (accessed 18 November 2023).

Stahl BC *et al.* (2023) A systematic review of artificial intelligence impact assessments. *Artificial Intelligence Review 56*, 12799–12831.

Stanford University Human-Centered Artificial Intelligence (2023) *Artificial Intelligence Index Report 2023*. Available at: https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AIIndex-Report_2023.pdf (accessed 25 November 2023).

Stoyanovich J and Khan FA. (2021) What is AI? Comic. Available at: https://dataresponsibly.github.io/we-are-ai/modules/what-is-ai/comic/ (accessed 18 November 2023).

Surveillance – Technology and Community – Safety Ordinance (2016) Municipal Code Chapter A40 -Santa Clara County, California, US. Available at: https://library.municode.com/ca/santa_clara_county/codes/code_of_ordinances?nodeId=TITAGEAD_DIVA40SUECCOAF (accessed 17 June 2022).

Surveillance Technology Ordinance (2018) Municipal Code Article 26.07 – Davis, California, US. Available at: https://library.qcode.us/lib/davis_ca/pub/municipal_code/item/chapter_26-article_26_07-26_07_010 (accessed 17 June 2022).

The New York City Council (2021) A local law to amend the administrative code of the city of New York, in relation to automated employment decision tools. Available at: https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4344524&GUID=B051915D-A9AC-451E-81F8-6596032FA3F9&Options=Advanced&Search (accessed 3 June 2022).

The White House (2023) FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence. The White House. Available at: https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-bidenissues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/ (accessed 27 November 2023).

Treasury Board of Canada (2023) Algorithmic Impact Assessment Tool. Available at: https://www.canada.ca/en/government/system/digital-government/digitalgovernment-innovations/responsible-use-ai/algorithmic-impact-assessment.html (accessed 26 November 2023).

U.S. Chamber of Commerce (2022) State-by-State Artificial Intelligence Legislation Tracker. Available at: https://www.uschamber.com/technology/state-by- state- artificial- intelligence-legislation-tracker (accessed 21 November 2023).

US President (2019) Maintaining American Leadership in Artificial Intelligence.

US President (2023) Executive order on the safe, secure, and trustworthy development and use of artificial intelligence. Available at: https://www.whitehouse.gov/briefingroom/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy development-and-use-of-artificial-intelligence/ (accessed 2 November 2023).

Verhulst S, **Young A and Sloane M** (2021) *The AI Localism Canvas*. Rochester, NY. Available at: https://papers.ssrn.com/abstract=3946503 (accessed 11 November 2023).

Wachter S (2022) *The Theory of Artificial Immutability: Protecting Algorithmic Groups Under Anti-Discrimination Law*. Rochester, NY. Available at: https://papers.ssrn.com/abstract=4099100 (accessed 11 September 2023).

Ward (2023) An act to add Section 11546.45.5 to the Government Code, relating to automated decision systems. Available at: https://custom.statenet.com/public/resources.cgi?id=ID:bill:CA2023000A302&ciq=ncsl&client_md=ff50b1aabd3b42d9e3 b51e6335aac82c&mode=current_text (accessed 5 October 2023).

Wegenverkeerswet (1994) Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. Available at: https://wetten.overheid.nl/BWBR0006622/2023-07-01 (accessed 26 November 2023).

Wyden and Booker (2019) Algorithmic Accountability Act of 2019. Available at: https://www.congress.gov/116/bills/s1108/BILLS-116s1108is.pdf.

Davis M (2020) Consumer privacy regulations: Considerations in the age of globalization and big data. In: Gianluca Cornetta, Abdellah Touhafi, and Gabriel-Miro Muntean (eds.), *Advances in Information Security, Privacy, and Ethics*. IGI Global Scientific Publishing: Hershey, Pennsylvania, United States of America, pp. 222–238.

Young M, **Katell M and Krafft PM** (2022) Confronting power and corporate capture at the FAccT conference. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. FAccT '22. New York, NY: Association for Computing Machinery, pp. 1375–1386. Available at: https://doi.org/10.1145/3531146.3533194 (accessed 24 March 2024).