If It Ain't Broke Don't Fix It: Steamboat Accidents and their Lessons for AI Governance

Bhargavi Ganesh, Stuart Anderson, and Shannon Vallor

Introduction

AI governance is treated by many scholars and practitioners as a proverbial Gordian knot, unsolvable in the absence of bold, unprecedented actions. If we look at the history of technology, however, we can see countless examples of these knots being methodically untied through innovative governance practices that promote transparency, safety, and accountability. In this paper, we use the example of steamboat regulation in the 1800's to challenge latent skepticism regarding the feasibility of governance of AI-driven systems[1].

First, we highlight the constructive[2] nature of US government responses to steamboat accidents, despite the limited governance resources available at the time. Second, we draw parallels between challenges to steamboat and AI governance and situate existing proposals for AI governance in relation to these past efforts. Finally, in noting some of the novel governance challenges posed by AI, we argue that maintaining a historical perspective helps us more precisely target these novelties when generating policy recommendations in our own interdisciplinary research group.

We start by framing our analysis through the lens of "responsibility gaps", first presented by Andreas Matthias (2004). Matthias builds on Aristotle's idea that both knowledge and control are required to attribute moral responsibility to an agent. He argues that because machine learning systems often do not enable sufficient human knowledge or control of system actions, it would be unjust in these cases to hold anyone responsible for the harms caused by the use of such technology (Matthias 2004: 176)· The inability to hold anyone responsible for a system action, while society still bears the cost of its consequences, represents what Matthias refers to as a "responsibility gap". While Matthias's original account was focused on knowledge and control, subsequent work by Coeckelbergh (2020) has also looked at the issue of "many hands and many things" (2052), generated by the increased number of stakeholders involved in bringing about the outcomes of AI-driven systems. Throughout the paper, we focus on the knowledge problem, which we refer to as "causal opacity", and the problem of "many hands", to demonstrate the ways in which similar responsibility gaps were effectively bridged by the constructive governance of steamboats.

After the first successful steamboat trip in 1807, steamboats were used regularly to transport people and goods, providing a major benefit to the US economy. But a series of deadly steamboat accidents quickly followed, and prior to 1860, more than 3,000 people were killed by more than 300 boiler explosions (Denault 1993: 99). Paradoxically, steamboat ridership continued to reach record highs, while people regularly voted for politicians who promised to

[1] The authors would like to acknowledge Burkhard Schafer, for his extensive feedback on this paper.

[2] Our use of the word "constructive" here does not refer to the legal idea of judicial construction, but rather constructive in terms of building on lessons learned from prior governance approaches, progressively devising new and better instruments to bridge governance gaps in ways that promote human flourishing.

regulate steamboat travel (Feenberg 2010: 40). Steamboat companies kept pushing their boats to travel faster, to justify the higher ticket prices they were charging consumers to offset high capital expenditures (Armstrong and Williams 2003: 169). These economic incentives did not prioritize the safety of passengers, operators, or members of the crew, and consequently, the US government decided to step in. Government intervention came despite the laissez-faire era of government at the time, in which concerns about state sovereignty and the view of regulation as stifling innovation meant that independent regulatory agencies and commissions had yet to be established.

Important parallels emerge when we compare the challenges posed by steamboilers and those posed by AI. Steamboiler technology, much like AI, suffered from causal opacity. It was difficult to understand how boiler explosions occurred, because the explosions often destroyed evidence, and forensic techniques were not yet developed enough to reconstruct the explosion from any evidence that was recoverable. Reporting on accidents typically came from the press and any surviving members of the crew. But press reports tended to be sensationalized, and members of the crew had incentives to present information in a way that would not implicate them as being blameworthy. Additionally, the use of steamboilers on boats was new and not well understood, and even scientists and operators differed in their assessment of the underlying technology (Maust 2012). Assigning responsibility for steamboat accidents was further complicated by the number of stakeholders involved in bringing about an outcome. There were “many hands” -- steamboat operators, steamboat managers, engineers, etc., as well as “many things” -- such as the vessel itself, the boiler, and navigation system, to name a few.

Federal government responses to steamboat accidents were constructive in addressing these challenges over time. Maust (2012) splits these interventions into four categories of governance ‘options’: information, mechanical, penalty, and regulatory. The information option involved collecting data on steamboat accidents and understanding the science behind bursting steam boilers[3]. With the help of think tanks such as the Franklin Institute, lawmakers conducted experiments and disseminated this information to the public (Brockmann 2002). The mechanical option involved subsidizing innovation around safety device engineering and design, and testing these safety devices[4]. The penalty option involved changing the liability standards and the rules of evidence to facilitate criminal prosecution and tort litigation, culminating in the passage of the 1838 Steamboat Act. Finally, the regulatory option involved passing the 1852 Steamboat Act, which set up examining and licensing of steamboat personnel, boiler testing, and set standards for steamboat construction and operation. Additionally, the Steamboat Act of 1852 set up the first ever regulatory agency, the Supervising Inspectors of Steamboats, which was involved with enforcing requirements, conducting inspections, overseeing accident investigations, and recommending changes to legislation[5].

We argue that there are parallel governance mechanisms available to us today which remain robust and promising options for governing AI, and provide a frame for analyzing potentially novel challenges. Furthermore, we observe that ahistorical narratives about AI’s supposedly

[3] Maust discusses the information option in detail in chapter 2 (starting on page 57)
[4] Maust discusses the mechanical option in chapters 3 and 4, starting on page 105.
[5] Maust summarizes the penalty and regulatory options on page ix.

intractable and unprecedented governance challenges are dangerously obscuring the creative power of technology governance to supplement existing governance mechanisms with innovative approaches when needed. Looking back at the governance of steamboats allows us to push back against false narratives about the tradeoff between governance and innovation, by demonstrating how governance can actually *enable* innovations in safety.

In our paper, we situate existing US, UK, and EU proposals for AI governance within Maust's framework, noting, for example, that government agencies, departments, and commissions have started to collect and disseminate information to aid standards setting. However, we also recognize that there are unique challenges posed by AI, which call for further constructive innovation in governance. The novel challenges of AI have precipitated the introduction of AI-specific governance proposals, such as the draft EU AI Act, which takes a risk-based approach (Veale and Borgesius 2021), and the US's proposed AI Bill of Rights, which takes a rights-based approach (White House 2021). In addition to analyzing these approaches, we conclude with some examples from our own interdisciplinary project on governing AI, that in seeking to meet these challenges, demonstrate the merits of a historically informed optimism about constructive and innovative AI governance.

Drawing upon approaches and literature in philosophy, science and technology studies, history, law, and policy, this paper proceeds in three parts. Part I provides a brief overview of previous work on responsibility gaps and articulates the usefulness of analyzing AI governance through this frame. Part II draws parallels between steamboat accidents and the harms posed by AI, while recognizing notable differences. Part III situates existing proposals for AI governance within Maust's taxonomy, expands on the parallel constructive governance approaches available to us today, and provides new policy recommendations to overcome the novel challenges posed by AI.

Part I: Contextualizing Responsibility Gaps Posed by AI

Ever since Andreas Matthias (2004) coined the term "responsibility gap", scholars at the intersection of AI ethics, policy, and governance have been using this concept to articulate novel challenges in governing automated decision-making systems. In response to Matthias's concept of a responsibility gap, some have argued that we should no longer deploy such systems (Sharkey 2010), while others have argued for the distribution of responsibility across human agents (Santoni de Sio and Mecacci 2021, Lima et al 2022). Still others have argued that there is no responsibility gap (Tigard 2021, Kohler et al 2017), or that responsibility itself is the wrong framing for understanding the disruptions posed by AI (Heinrichs 2022).

De Sio and Meccacci (2021) point out four types of responsibility gaps introduced by AI-based systems: the culpability gap (lack of a human agent to blame/hold culpable), the moral accountability gap (inability to understand why the system brought about an outcome), the public accountability gap (lack of answers provided to the public for a given outcome), and the active responsibility gap (lack of forward-looking responsibility to ensure that designers and users act according to moral obligations) (1059). Kohler et al (2017) deny the existence of a culpability gap (which they refer to as an accountability gap) on the basis that there can always be someone for whom blame is retrospectively assigned for a given outcome. Focusing on the issue of forward-looking responsibility, Lima et al (2022) follow Coeckelbergh (2020)'s framing of the importance of agents of responsibility providing answers to "patients", or those who are affected by the actions of a responsible agent (2061), and argue for a strict notion of forward-looking responsibility in which regulators hold designers prospectively accountable for the outcomes of AI-based systems.

Even if responsibility is prospectively assigned, however, there is still the issue of "many hands and many things" (Coekelbergh 2020: 2052). To address this issue, Sven Nyholm (2018) argues for hierarchical human-robot collaborative teams, in which we can locate the responsibility for a given event based on the degree or type of agency that the human agent has in that situation, while De Sio and Meccacci (2021) recommend "designing technical systems for meaningful human control", and mapping responsible stakeholders to specific responsibility gaps (1075). Similarly, Heinrichs (2022), who denies the usefulness of responsibility as a concept altogether in this context, argues that the introduction of AI necessitates "identifying different moral relations" between a complex set of actors (8).

In our analysis of the literature on responsibility gaps, we can see that even those who deny the existence of such gaps still acknowledge that AI-driven systems pose challenges to the distribution of moral and legal responsibility. In practice, these ambiguities have given individual stakeholders the opportunity to deflect responsibility for the outcomes of these systems, to the detriment of those affected and society at large. In this paper, we do not wade further into the debate about whether responsibility gaps are a matter of reality or perception. We argue that at the very least, the lack of clearly articulated responsibilities for the outcomes of AI-based systems creates an *apparent* gap between the outcomes of these technologies and the human agents involved in their design, deployment, and operation, and that the presence of this apparent gap is sufficient to justify analyzing policy and governance approaches from this lens. To that end, our analysis of steamboat accidents in the 1800's highlights the federal government's actions in bringing down the number of steamboat accidents to demonstrate the way that constructive governance approaches can counter the responsibility deflections that arise in the presence of the ambiguities in responsibility attribution.

Part II: Turning to the Case Study of Steamboat Accidents in the 1800's

Part II is divided into four subsections. In subsection A, we motivate the use of the steamboat case study. In subsection B, we provide some historical background on the rise of steamboat accidents and resulting public calls for government intervention. In subsection C, we detail US Federal government responses to steamboat accidents. In subsection D, we draw parallels between the challenges posed by steamboats and the challenges posed by AI.

A. Motivating the Use of the Steamboat Case Study

Many scholars discuss the disruptive impact of AI as if it is an entirely new phenomenon. An ahistorical perspective has contributed to "AI hype", which has been characterized by overestimation of the technical functionality of AI applications (Raji et al 2022), and also fueled a narrative that the novel complexities of AI render the governance mechanisms we have built over centuries hopelessly ineffective. Additionally, as Gifford (2018) notes in his analysis of the impact of the industrial revolution on tort law, even those who *do* analyze the history of technology, such as legal scholars and many historians, tend to "ignore technology's role in causing unwanted and deleterious consequences" (75).

Legal and policy proposals related to AI have recently not only understated harms, but also overstated unlikely harms, thereby distracting policy aims[6]. For example, US Senate proposals discuss "AI systems gone awry", and earlier versions of the EU AI Act treated AI as being part of some sort of sci-fi fantasy (Floridi 2021: 218). Grounding policy recommendations in the history of technology is therefore important in helping researchers and policymakers appropriately calibrate the risks that these technologies pose and remain optimistic about the potential of governance to solve the issues at hand. To that end, in Part III, we situate existing proposals for AI governance within a framework informed by historical context.

Of course, steamboat accidents are not the only historical example of governance in response to the introduction of new technology. Our motivation for focusing on steamboats in this paper is two-fold. For one, steamboat regulation represents the first independent national regulatory effort and led to the creation of the first Federal Agency in US history (Burke 1966: 3). The fact that constructive modes of governance actually *promoted* innovation in safety, provides a powerful counter to the idea that there is a tradeoff between innovation and regulation-- an argument which is often used to discourage the regulation of AI. Second, the passing of steamboat regulation required many instances of trial and error before it was ultimately successful at improving safety. Though it took 30 years before the final version of the Steamboat Act was passed, the "soft" measures of governance leading up to that point enabled safer steamboat design, and served as an important building block for future regulatory measures. We remain similarly optimistic that the current measures being undertaken for AI governance can lead the way to constructive and innovative regulatory measures.

B. A Brief History of Steamboat Accidents in the US

In 1769, James Watt invented the first steam engine, with the help of manufacturer Matthew Boulton. Watt and Boulton designed the steam engine with low-pressure steam, and even early on, they opposed the use of high-pressure steam on the grounds that the danger of explosion was not worth the efficiency gains (Leveson 1992). Robert Livingston and Robert Fulton's first commercially successful steamboat, *The Claremont*, stayed true to Watt and Boulton's steam engine design. By 1800, however, Watt and Boulton's patent expired, and Oliver Evans built the first high-pressure steam engine in the US. Even though Livingston and Fulton had a monopoly agreement with the state, Henry Shreve successfully broke the monopoly and used high-pressure steam engines for the boats he operated along the Mississippi River (Hunter 1943).

High-pressure steamboats soon became ubiquitous and were lauded for their role in increasing economic growth through their efficient transport of goods and expansion of personal travel. The expansion of personal travel was particularly exciting because prior to the steamboat era, riverboats and keelboats lacked the size or structure to carry paying passengers (Buchanan 1967: 9). In the "golden age of the steamboat", increasingly ornate steamboats included private decks for higher-class travelers, and provided what was at the time considered a luxurious form of transportation (Burton et al, n.d).

[6] For a full discussion of scientifically inaccurate pop culture understandings of technology making their way into proposed regulations on autonomous vehicles, see McLachlan et al (2022).

Despite the excitement around steamboat travel, the steam boiler explosions Watt had feared eventually came to pass, and from 1811 to 1851, according to the US Army Corps of Engineers, 21% of river accidents occurred due to these explosions. While there were other causes for steamboat accidents such as collisions with other boats, obstructions in rivers, and vessel fires, steam boiler explosions caused the most fear and anxiety among the general public, and these explosions were regarded as the largest man-made disasters other than war (Brockmann 2002: 5). Press coverage of boiler explosions included gory images and harrowing accounts. Emotional pleas for action to curb these accidents became more and more pronounced with time, and the reduction of steamboat accidents quickly became a priority in both parties' political platforms (Brockmann 2002: 43).

Of course, boat travel always came with risks, even before the age of steamboats. What made these disasters different, however, was not just the graphic nature of the injuries and deaths, but also the fact that the deaths now included passengers. Still, some people viewed the explosions as an acceptable cost of innovation (Leveson 1992: 3). In fact, Mark Twain famously romanticized steamboat travel on the Mississippi (Buchanan 1967: 5), even though he had lost his own brother to a steam boiler explosion (Brockmann 2002: 129).

The press and politicians tended to focus on the "reckless" behavior of steamboat operators and crew, and tended to indemnify the technology itself, despite the widely publicized dangers of high-pressure steam. The prevailing view of the public and politicians alike was that "operators lacked either the knowledge or the care necessary to control such powerful agents" (Rice 1963: 117). As Rice (1963) and Buchanan (1967) note, steamboat operators and crew members were a convenient scapegoat because they tended to be of lower socioeconomic class, and comprised of a mix of recent European immigrants, who were generally viewed in a prejudiced manner at the time, as well as free Black and slave laborers. Ironically, members of the boat crew tended to be the most affected by the boiler explosions--- it was not uncommon for more than a quarter of the slave workers to die on an exploded steamboat (Buchanan 1967: 58-59).

Some criticisms of steamboat operation were legitimate—namely the practice of "racing" steamboats, and pressing down on safety valves to make the boats go faster (Brockmann 2002: 14). While some of these practices could be attributed to the "machismo" of individual crewmembers, in general crewmembers tended to face pressure from steamboat owners, who pushed for boats to move as fast as possible, to maintain their competitive advantage. Nevertheless, the prejudice against steamboat operators motivated the federal government to pursue the penalty option, which we discuss in more detail in subsection C.

C. Governance Responses to Steamboat Accidents: Maust's taxonomy

Over the span of thirty years, the US government took a variety of different approaches to ensure the safety of steamboats. Maust (2012) places these approaches into four different categories of governance "options". The information and mechanical options involved disseminating information about steamboat accidents and steamboiler technology, and encouraging innovation in safety devices through government subsidies and evaluations. The penalty and regulatory options were reflected in the passing of the 1838 and 1852 Steamboat Acts. The penalty option emphasized the criminal negligence of steamboat operators and owners, whereas the regulatory option expanded administrative oversight over both the technical artefacts and the humans involved. In this subsection, we analyze each of these options and discuss their relative effectiveness or ineffectiveness from the standpoint of steamboat safety.

Information Option

According to Maust's archival research, Congress published 80 documents from 1824 to 1852, with the goal of improving scientific literacy about steamboats amongst not only the general public, but also engineers, who tended to have varying degrees of training[7]. The published documents provided information about the causes of steamboat disasters, as well as technical advice on how to safely build and operate steam vessels and steam engines. In conducting investigations about steam boiler safety, the government teamed up with the Franklin Institute, a private research institute made up of experienced mechanics and academics. Using government funds, the Franklin Institute conducted experiments over a span of five years to understand how steam boilers worked, and the conditions that caused them to fail (Burke 1966: 8).

While the information option likely succeeded in empowering at least some designers and manufacturers to build safer systems, Maust argues that one of the main drawbacks was that the government struggled to differentiate between false claims and scientifically grounded evidence. There was a so-called "avalanche of pseudoscience" to contend with (Maust 2012: 76)[8]. Additionally, the general public, and even many engineers and steamboat operators, either did not find the documents interesting enough to read, or disagreed with their conclusions, often preferring to believe common myths instead[9]. McLachlan et al (2022) present an example of one such myth, outlined by Burke (1966), in which the public believed that copper boilers were safer than iron boilers, because the press reported that the boiler involved in a major accident was made of iron. In reality, the copper boilers tended, at first, to be used with lower pressure steam, and it was the higher-pressure steam that had caused the accident. Changing public perception proved impossible, however, despite the efforts undertaken under the information option.

Mechanical Option

Using the "mechanical option", the US government provided incentives for the development of safety-related innovation through multiple means. For one, inventors were allowed to petition for aid, giving them an opportunity to experiment and build safety devices and other inventions to improve the safety of steamboats[10]. Additionally, in 1834, Congress appropriated funds to the Navy to test steam boiler safety equipment. The Navy picked the safety instruments to test on an ad-hoc basis and ultimately recommended a set of safety devices and related products[11]. The

[7] Maust discusses his archival research methodology, as well as the 80 documents on page xv. He also discusses the impact of these documents on the development of best practices on page 100.

[8] Maust cites *The Papers of Joseph Henry*, a primary source from 1844. Joseph Henry was a scientist who was critical of the willful ignorance of those who were overconfident in their ability to master nature through technical applications.

[9] Maust discusses the issue of quality control in government reports in Chapter 2, starting on page 46.

[10] Maust discusses the process of petitioning for aid and government subsidies more broadly in Chapter 3, starting on page 105.

[11] Maust discusses the Navy's interest and intervention from pages 106-123.

government was also tempted to mandate specific safety features on steamboats, such as the requirement of using iron chains and rods, which made its way into the 1838 Steamboat Act. However, in later iterations of the steamboat regulation, Congress scrapped the requirement because steamboat owners and operators found the requirement counterproductive, particularly in storms (Maust 2012: 124-125).

Over time, through government intervention and significant work on the part of inventors of safety devices, the safety profile of steamboat technology improved considerably. However, steamboat accidents persisted. There were still operational challenges that needed to be addressed, such as the need for safety protocols to ensure that crew members and staff did not panic during emergencies. Instead of viewing these challenges as part of a larger system of human-machine interactions that might be addressed through system design and regulation, some legislators viewed the technical and operational challenges as being inherently separate and in need of distinct interventions. This led them to pursue the penalty option, which we detail in the next section.

Penalty Option

The penalty option of governance refers to the use of private law or criminal sanctions *after* steamboat accidents occurred. Prior to the writing of the 1838 Steamboat Act, there were already a number of court-imposed sanctions in place at the state level, including fines or imprisonment in the criminal case, and monetary awards in the case of civil/tort liability. In designing a federal law, the government had the following options; more rigorous enforcement of existing laws, codifications of existing laws to facilitate enforcement, such as changes in the burdens of proof, or the creation of entirely new offences and private law actions. In the Steamboat Act of 1838, the government ultimately opted for a combination of these, creating a new criminal law "manslaughter provision", which also stipulated procedurally that the steamboat accident alone served as *prima facie* evidence of negligence, making it possible for steamboat owners and/or crews to be charged with manslaughter, serving up to ten years of hard labor. The law also required that if a steamboat owner or crewmember were charged, they would have to prove their innocence (Maust 2012: 211-212), by showing, for example, that a third party had tampered with the steam boiler.

In the 1850s, there was a broad recognition that the 1838 law had not been effective in bringing down the number of steamboat accidents (Brockmann 2002)[12]. Though there were provisions included which suggested that boilers needed to be inspected, and that the vessel needed to be in a "seaworthy" condition, the law remained vague about exactly how this should be ensured (Maust 2012: 211). Inspections were perfunctory at best, and bribery was not uncommon (Hunter 1949: 533-534). The narrow focus on the criminal liability of owners and crewmembers also depended on the often false assumption that these individuals had enough knowledge or control to change a given outcome. In reality, explosions could rarely be explained by human fault alone-- typically they were the result of the interaction between negligent human operators and bad design and manufacturing choices that allowed small human mistakes to have grave consequences. The regulatory option discussed below, in which operators, engineers, and

[12] Brockmann discusses the failure of the 1838 Steamboat Law in Chapter 5.

inspectors were all certified, along with the technical artefacts themselves, offered a more comprehensive strategy that was in line with this reality.

Regulatory Option

The rewriting of the 1838 Steamboat Act enjoyed broad support from a wide swath of civil society groups, including immigrant rights organizations and steamboat owners, who pushed back against the manslaughter provision (Maust 2012: 218-220). However, as Mashaw (2008) notes, despite this pushback, there were some who argued that a new law merely needed "enhanced civil liability", and "more stringent safety requirements" (1637). Instead, the new steamboat law removed the manslaughter provision and took a more innovative route, focused on designing preventative regulatory measures, administrative enforcement provisions, and administrative remedies (Mashaw 2008: 1641).

The 1852 Steamboat Act created the first independent regulatory commission, the Supervising Inspectors of Steamboats, which was made up of "sixty inspectors, who would be stationed at ports throughout the country under the direction of nine supervising inspectors appointed by the president and confirmed by the Senate" (Maust 2012: 18). The regulatory body was tasked with enforcing standards for vessel and boiler design and construction, testing boilers, checking for safety equipment onboard, and issuing licenses for the vessels to allow them to operate. Inspections were conducted on an ongoing basis, could occur without prior notice, and could be the basis for revoking vessel licenses (Mashaw 2008: 1641). The new law also provided detailed information about the qualifications needed to become an inspector, added a provision that inspectors would need to be paid government salaries, and contained detailed subsections on the duties of inspectors and the penalties for failing to meet these duties (Maust 2012: 228). The Supervising Inspector of Steamboats subsumed some of Congress's previous work as part of the mechanical option, by continuing to encourage inventions of new safety mechanisms and helping steamboat owners comply with the technical requirements posed by the 1852 Steamboat Act (Maust 2012: 246). The Steamboat Act was also inspired by regulatory measures in Britain and France, and consequently included criteria on the professional licensing of engineers (Maust 2012: 224), and allocated federal funds to save victims of shipwrecks, aid navigation, and remove obstructions from waterways (Means 1987).

The 1852 Steamboat Legislation was a culmination of years of efforts on the federal level, which included 62 bills and joint resolutions that were introduced from 1824 to 1860 (Maust 2012: 255). Maust's account presents steamboat regulation as a success of both constructive and innovative governance:

> "Because they had examples where Congress had already used its power to regulate commerce for sailing vessels, the regulations in the 1852 Steamboat Act were innovative in instituting more extensive requirements than other passenger laws, along with a new bureaucracy, rather than putting in place an entirely new form of national government intervention". (Maust 2012: 248)

In drawing parallels between steamboats and AI in subsection D, we are implicitly optimistic that parallel forms of constructive and innovative governance are available to us to solve our modern technological governance challenges.

D. Drawing Parallels between Steamboats and AI

The issue of causal opacity, in which the press, members of the public, legislators, operators of the technology, and engineers themselves often did not understand how steam technology worked is one that parallels our current challenges with AI. There are a number of reasons for causal opacity in AI. For one, opacity arises from the use of models whose decision-making is based on complex relationships between the target variable and input variables; relationships which are often not interpretable by human beings. An added layer of opacity is created by the fact that machine learning models tend to be trained on proprietary datasets which cannot be scrutinized by independent researchers. This can be due to commercial reasons, or data privacy laws such as HIPAA laws in the US. It is therefore hard for independent researchers or investigative reporters to scrutinize these models (Raji et al 2020, Costanza-Chock et al 2022).

The press continues to be an important vehicle for surfacing potential harms generated by new technologies. For example, VICE reported that a model used on X-ray images inadvertently detected race categories (Feathers 2021). This example is emblematic of current reporting on AI, which follows a similar pattern to reporting on steamboats-- reporters surface a potential harm, note that the researchers themselves do not know how it came about, and this makes headlines everywhere, until the next potential harm is surfaced. Relying on the press to surface harms has limitations, however, especially since the subtlety of harms in AI means that only a subset of them can be identified and reported on. AI is similar to steamboats in this way as well—crews and operators had been dying on riverboats and keelboats for years, but the press only started reporting on steamboats after a series of graphic boiler explosions led to the death of passengers.

Additionally, reporting on AI tends to spin causal opacity in sensational ways that can misleadingly overstate AI's capabilities. Just as reporters described steam technology as "mysterious" and "powerful", reporters use similar epithets to describe AI, and several media outlets reported on an engineer who was convinced that an AI-based system is sentient, despite members of the research community vehemently denying that this claim had scientific basis (Bender 2022). Causal opacity has also been taken advantage of by opportunistic and misleading actors. The same "avalanche of pseudoscience" decried by scientists in the 1800's exists around AI as well. For example, several companies have sold devices with the promise of using AI for emotion detection, but researchers have argued that this too is based on pseudoscience (Barrett 2019).

In addition to the causal opacity problem, the "many hands and many things" problem is similar when we compare AI-based technologies to steamboats. However, now there are now many *more* hands and many *more* things than there were in the steamboat era. And the fact that AI-based systems can make decisions, for which the systems themselves cannot be held responsible, makes it that much easier for individual actors to deflect responsibility onto one of the other "hands" or "things". We can demonstrate this using the example of a hypothetical hiring algorithm. Let's say a company decides that the deluge of applicants they are receiving necessitates the use of an AI-based system to decide which candidates should be shortlisted. They enlist the help of a company like Ideal, which provides them with a customized machine-learning based solution for screening resumes. The AI-based system screens out candidates, and ultimately chooses an applicant. In this entire process, we can see there are many hands involved—within the company creating the system there are the designers, product managers, testers, etc., as well as those who collect and manage the resume data that is being used to train the models. The company using the system also has many hands, including HR personnel and managers involved in procuring and using the technology. There are also many things—the resumes collected for training/testing, the resumes used by the organization that is hiring, the AI-based system, the models underpinning it, etc. If there is an adverse outcome publicized, such as the unfair screening out of an applicant based on a spurious feature, like the length of their name, it is possible to see how the company using the hiring algorithm could deflect responsibility to the company designing the algorithm, and that company could in turn justify the outcome based on data constraints, etc.

In response to the issues of causal opacity and many hands, there have been a number of technical and process-driven governance approaches proposed by researchers. Some scholars have proposed requiring that models be made in an interpretable manner (Rudin 2019), whereas another area of research has emerged around post-hoc explanations of model behavior (Ribeiro et al 2016, Lundberg and Lee 2017). Others have focused on process-level transparency, and have called for designers to create model cards (Mitchell et al 2018), or datasheets (Gebru et al 2018). Still others have called for more robust evaluation methods, and checklists to prevent common errors and pitfalls (Kapoor and Narayanan 2022). Researchers have also proposed the conducting of independent audits (Raji et al 2020) or impact assessments (Selbst 2021), and licensing for AI models so that they can only be used for their designed purpose (Contractor et al 2022).

Many of these proposals mirror the steamboat era—research in model transparency, interpretability, and evaluation are in some ways similar to safety valves, independent audits/impact assessments are akin to steamboat inspections, and AI model licensing is not too different from the coastal licenses required to operate boats. Much like the professional licensing of steamboat engineers, there has also been some discussion about requiring professional licensing of the data scientists and/or machine learning engineers who design and develop machine learning models. At the moment, some of these efforts are being conducted by the private sector, i.e. model cards are being implemented at Hugging Face, IBM and other companies have developed tools for "explainable AI", auditing criteria is being developed by companies like ForHumanity, and independent audits are offered as a service by companies like Parity AI and ORCAA (Costanza-Chock et al 2022). Far from workable AI governance options being unavailable, what we really have are a surplus of promising ideas that require coordination, testing, implementation, evaluation and revision through the same sort of active learning process involved in constructing and adapting effective steamboat governance regimes.

Yet we note that there are certain characteristics of AI that are inherently new. First, the growing scale, speed and autonomy of AI-based systems, means that they not only take over certain agential powers from human agents, but also that they can acquire new agential powers (that is, the ability to perform unprecedented actions), from which new classes of potential harms emerge (Vallor and Ganesh forthcoming). For example, the potential to track, analyze and even predict the movements of an entire city's population in real-time was simply unrealizable by human agents previously, but becomes a new agential power with AI. With it comes the opportunity for novel privacy harms and political abuses. Second, while steamboat technology was constrained to power generation, AI is being used in a far wider range of domains; it is increasingly a "general-purpose" technology. The application areas are growing at a pace that threatens to overwhelm governing bodies tasked with disseminating information and guiding best practices on AI's use. Third, the reliance of these systems on datasets that can be continuously updated presents new challenges for external examiners who aim to accurately characterize model performance at a given snapshot of time, let alone assess potential future harms. While some have pointed to the FDA's guidance proposed predetermined change control plan[13] as a potential framework for addressing these challenges, operational challenges to implementing robust post-market surveillance still remain.

In Part III, we first acknowledge the governance mechanisms being pursued today which parallel those pursued during the steamboat era, and then discuss proposed approaches such as the EU's AI Act and the US AI Bill of Rights, alongside our own policy suggestions for addressing the novel governance challenges posed by AI.

Part III: Situating Existing and Proposed AI Governance Approaches within Historical Context

As demonstrated in the previous sections, Maust's taxonomy of governance responses to steamboat accidents presents a useful frame for analyzing AI governance. An exhaustive survey of current governance approaches is outside the scope of this paper. However, we provide some examples to illustrate the character of current efforts, and provide a starting point for historically informed critical evaluation of both current and proposed efforts.

In the US, the governance of AI appears to be a priority of the White House's Office of Science and Technology Policy[14], and at least nine[15] federal agencies have taken overt steps towards understanding the impact of AI on their respective area of governance. In the UK, the government created the Center for Data, Ethics and Innovation (CDEI)[16], which coordinates efforts on AI regulation with different departments and ministries in the UK. The EU, on the other hand, conducts its AI regulation efforts under the umbrella of the EU digital strategy. In this section, we discuss current governance approaches classified broadly under the information and mechanical options, briefly acknowledge some of the challenges to legal liability (tort/contract law, as well as regulatory enforcement) posed by AI, and then focus our remaining discussion on analyzing proposed ex-ante measures that would be enforced by regulatory bodies.

Combatting Causal Opacity Through the Information and Mechanical Options

Both the information option and the mechanical option are currently being pursued by the US federal government, as well as the UK government and the EU[17]. In the context of AI, we take

[13] See FDA discussion paper: https://www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf

[14] https://www.whitehouse.gov/ostp/ostps-teams/science-and-society/

[15] Referring to efforts being undertaken at FDA, CFPB, FTC, DOL, EEOC, FHFA, Federal Deposit FDIC, Office OCC, as well as NIST, which is a subsidiary of the DOC.

[16] The CDEI is an advising body recently created within the Department of Culture, Media, and Sport (DCMS)

[17] For simplicity, we limit our discussion to efforts taken by the UK and EU broadly and do not include specific efforts being taken by individual member states in the EU, and separate efforts

the information option to mean the collection and public dissemination of expert information about AI risks and opportunities, and the mechanical option to mean the distribution of research funding, establishing of partnerships focused on testing or evaluating systems, and emerging government recommendations on best practices for design and implementation of AI-based systems.

Governments have undertaken the information option in a variety of ways. US government agencies have historically used Requests for Information (RFIs) to gather information before issuing regulatory guidance, and they have used this approach to understand AI policy as well[18]. Similarly, in the UK, the CDEI has generated calls for evidence, on topics such as bias in algorithmic decision-making (Centre for Data Ethics and Innovation 2019). The US, UK, and EU, also convene conferences/workshops inviting expert input, and publish informational reports on their government websites on a wide range of topics, including but not limited to the use of AI in healthcare, genomics, and financial services, as well as technical considerations such as explainability, model evaluation, and cybersecurity risks.

Research funding is a major part of the mechanical option for AI governance. Recent AI-related research funding by the National Science Foundation (NSF) in the US has focused on the societal impacts of AI. In the UK, UK Research and Innovation (UKRI) has provided £33M in funding for interdisciplinary work that addresses social and technical challenges in the design, regulation, and operation of trustworthy autonomous systems[19], and in the EU, Horizon Europe has funded projects related to trustworthiness in AI[20]. In partnerships that resemble the partnership between the Franklin Institute and the US government during the steamboat era, the UK research council provides funding to both the Alan Turing Institute and Ada Lovelace Institute. The Turing and Ada Lovelace Institutes are involved in both the information and mechanical options of governance, by producing reports that inform the public[21], and actively shaping AI practice through the development of criteria for evaluation/sociotechnical audits of systems[22]. While the US government has not formally funded or partnered with one particular institution, NIST has worked with academics to design and conduct evaluations of AI systems, and produced technical reports after each evaluation[23].

The information and mechanical options being undertaken globally already show promise in providing different perspectives, engaging a wide set of stakeholder groups, and tackling more

being taken by countries like Scotland within the UK. We do not discuss efforts being taken by individual U.S states.

[18] For example, the Department of Treasury, Federal Reserve Board of Governors, FDIC, CFPB, and National Credit Union Administration issued a joint RFI on financial institutions' use of AI.

[19] See https://www.ukri.org/news/new-trustworthy-autonomous-systems-projects-launched/.

[20] See https://ec.europa.eu/info/research-and-innovation/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe_en.

[21] See Brennan (2019) and Ostmann and Dorobantu (2021), for example.

[22] See Ada Lovelace Institute (2022) and Centre for Data Ethics and Innovation (2021), for example.

[23] Evaluations described here: https://www.nist.gov/system/files/documents/2022/03/17/AI-RMF-1stdraft.pdf

ambitious societal objectives than were even envisioned in the era of steamboats. There are opportunities for even more creative governance approaches as well. For example, governments could voluntarily request companies to share model cards of their proposed implementations, and these could be reviewed by governing bodies and released to the public to promote transparency. Governments may also consider surfacing evidence of harms through complaint databases, and further community outreach efforts.

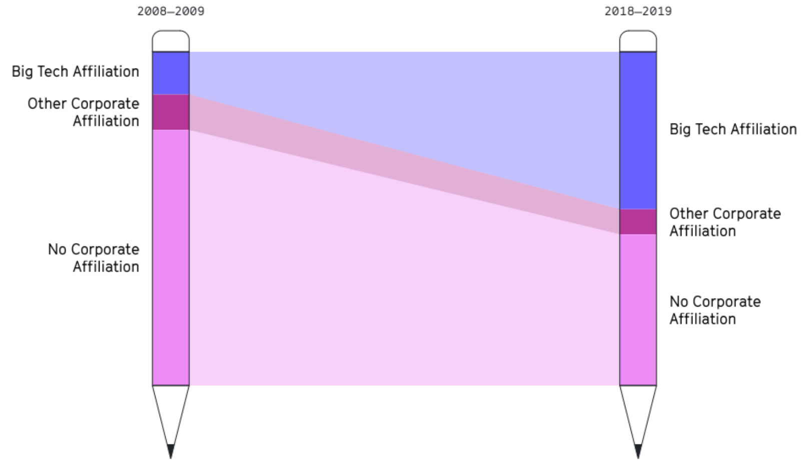
However, it remains to be seen how effective the information and mechanical options of AI governance will be. For one, much like the steamboat era, it is hard to imagine that there is much demand from the general public for the reports produced by the government. Our current digital age presents even more of an information overload, which is likely crowding out official government sources. Similarly, it is unlikely that data scientists and other professionals involved in the design and deployment of AI look to government sources for any technical understanding. Second, some government reports demonstrate a temptation to produce reports outlining best practices, but much like the steamboat case, these efforts could backfire if they are undertaken too early. For example, in the UK, the ICO produced a report which prescribes best practices for explaining decisions made by AI, even though best practices not yet emerged amongst scholars in this area. Additionally, the report does not cite any specific literature or sources for the recommendations made. In contrast, NIST's draft report on an AI Risk Management Framework provides a clear overview of the challenges of using explainability methods, while situating them within a broader framework of other evaluation techniques that can be used to improve the risk profile of a given AI implementation. It is important that governments remain nuanced in their recommendations, to avoid creating mere compliance box-ticking exercises, which could in some cases make systems less safe, particularly if they engender unwarranted trust in system capabilities.

There are also significant new challenges emerging, such as the issue of industry capture in academia (Whittaker 2021). Of course, private industries have always had an influence on policymaking, and working with industry to understand this emerging area is not inherently negative. However, academics, like the scientists and engineers at the Franklin Institute who were consulted in the steamboat era, have tended to, for the most part, provide the independent technical expertise needed to drive the information and mechanical options. In the area of machine learning research, there has been a significant blurring of lines between academia and industry due to the monopolies that large tech companies have over data and data infrastructure. Additionally, these same companies fund major university projects and PhD fellowships. This impact is clearly borne out by Birhane et al (2022), in which the authors analyze the 100 most cited papers from the top machine learning conferences, ICML and NeurIPS, and find that big tech author affiliations went up from 11% to 47% in the span of 10 years. The data visualization in Figure 1 below clearly demonstrates this dramatic increase.

Importantly, clear conflicts of interest emerge if those who are advancing the technical field are affiliated with companies which are unlikely to want to disclose all the potential harms of their models, due to their desire to avoid regulation and protect their reputation. This likely diminishes the effectiveness of the government in ensuring quality control in the dissemination of information, particularly in ensuring the publicizing of *all* known AI risks. Conflicts of interest also likely diminish the effectiveness of research funding in promoting genuinely responsible practices in AI. Additionally, the market power of these companies is so large that it is hard to imagine governments being able to offer a large enough incentive for competitive innovation in research around improving the safety of these systems, particularly if doing so would come at a cost to the business bottom line.

These challenges are not insurmountable, however, and could be addressed through governments providing more university funding for data and data infrastructure, funding projects that creatively identify harms and evaluate existing deployments of technology, and engaging more with scholars who do not have compromising industry affiliations. An even more ambitious and creative option is to require that industry directly subsidize regulatory activity as compensation for the taxpayer- funded research and higher education upon which AI companies depend; for example, taxation policies could redirect profits from industry to subsidize more competitive salaries for AI experts and regulators in the public sector, or guarantee the cancellation of student loans for those who study AI policy and regulation and complete a term in regulatory service.

Figure 1: Author Affiliations in highly cited papers over 10 years



ⓘ This shows that the corporate affiliations of authors of the 100 most highly cited papers from two major conferences increased significantly from 2008–2019.

The Values Encoded in Machine Learning Research, Abeba Birhane, Pratyusha Kalluri, Dallas Card, William Agnew, Ravit Dotan, Michelle Bao, 2021

Source: Mozilla Foundation Internet Health Report, 2022

Addressing "many hands": Legal Liability and Preventative Regulatory Measures

Since the steamboat era, tort and contract liability have expanded in scope and reach, and have greater relevance to the discussion on AI liability than the criminal negligence standard. Despite innovations in tort and contract liability, however, the introduction of AI-based systems in a variety of contexts has challenged these existing frameworks, and provided opportunities for companies to deflect legal responsibility, such as in the case of vehicle injury lawsuits related to Amazon's flex app (Soper 2021). While there is some optimism that product liability legislations can be amended to accommodate the unique challenges posed by the "many hands" involved in design and deployment of AI (Villasenor 2019), the UK's Office of Product Safety and Standards (OPSS) notes that existing product liability regulations would face challenges if applied to AI due to "the blurring of the lines between products and services; the increasing ability for consumer products to cause immaterial as well as material harm; the increasing complexity of supply chains for consumer products; and issues related to built-in obsolescence and maintenance throughout a product's lifecycle"[24]. The EU has also noted that the lack of a specific liability regime for AI makes it challenging to determine when it would be best to apply product liability, versus general torts, or contract liability[25].

Additionally, there are questions regarding the legal liability regime that should be used for regulatory enforcement of AI-specific regulations such as proposed regulations requiring algorithmic impact assessments. Yew and Hadfield-Menell (2022), like Selbst (2021), argue that federal agencies in the US could use penalty default regimes to incentivize prospective harm mitigation on the part of designers of AI-based systems. A discussion about the ideal liability regime for AI is outside the scope of this paper. While we recognize the importance of creative governance in liability regimes for the purposes of civil litigation, as well as enforcement of regulatory standards, in this section we focus on what we can learn from the steamboat analogy as we consider what evidence companies should be required to provide to regulators, and which regulators would be best positioned to analyze this evidence.

As AI has started to be deployed in a wide range of industry sectors, the US government has begun issuing guidance on how the use of AI in these sectors would impact compliance with existing laws, i.e, the EEOC's guidance on complying with the American Disabilities Act. Comprehensive AI-specific legislation has not yet been enacted, but legislations such as the Algorithmic Accountability Act, and the Algorithmic Fairness Act of 2020, have been introduced in Congress. Both legislations authorize the Federal Trade Commission to enforce laws that would require companies to assess the potential negative societal impacts of their products. However, as we have demonstrated, there are already a number of agencies beginning to develop guidance on the use of AI-based systems. This suggests that a sectoral approach to AI governance, in which these agencies are supported through congressional appropriations, is a more constructive way forward. The proposed AI Bill of Rights could be useful to the extent that it clearly articulates the ways in which AI-based technologies can infringe upon the fundamental rights that have been established in the US constitution. This articulation would then help agencies pursue the relevant enforcement mechanisms to ensure that a given right is protected.

The Algorithmic Accountability Act also suggests the use of algorithmic impact assessments. Impact assessments have traditionally occurred in the environmental policy space, where they were used to scope out the potential hazards of a given project. Scholars such as Selbst (2021) suggest that an algorithmic impact assessment could be used in a similar manner by companies to demonstrate they have considered the potential harms associated with a proposed implementation. Relatedly, Raji et al (2020) argue that algorithmic audits should be used to determine the compliance of companies to internal or external standards. Audits have

[24] See OPSS report: https://www.gov.uk/government/publications/study-on-the-impact-of-artificial-intelligence-on-product-safety

[25] See EU report: https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/JURI/DV/2020/01-09/AI-report_EN.pdf

traditionally been used in the financial sector for the purpose of regular monitoring of performance of a company each year. Both those who propose algorithmic audits and algorithmic impact assessments seem to generally view them as evaluation tools pre-deployment, and on a regular basis post-deployment.

As mentioned earlier, while there are already a number of companies conducting algorithmic audits, there are no established audit standards, and due to the lack of regulatory backing, these audits are generally done on an ad-hoc or voluntary basis. As the steamboat case study shows us, it is important that inspectors, or in this case auditors, are independent and incentivized to conduct audits in a way that is fair and accurate. In revisions of the Steamboat Act, legislators focused on ensuring that inspectors received government salaries, were certified, and faced penalties for failing to do their job in an independent manner. More recent cases, such as the Enron scandal, have also highlighted the importance of ensuring that auditors are held accountable. The importance of regulatory oversight is also emphasized by Raji et al (2022), who provide a template for regulators carrying out algorithmic audits. Of course, there are definite risks of industry capture even when government agencies are in charge of inspecting systems, as we saw in the case of the Federal Aviation Administration (FAA)'s relationship with Boeing during the 737-MAX certification. However, these are known challenges that governments have the experience and tools to address.

In our interdisciplinary research group, we have studied the types of gaps that may emerge from new agential powers resulting from introducing AI in areas that used to be driven completely by human decision-making (Vallor and Ganesh forthcoming). While it is impossible to hold a system itself accountable for the outcomes of machine learning decisions, the fact remains that AI-based systems can generate outcomes that have little or no precedent in human action and governance. The responsibility gap, as we define it then, comes about because AI not only takes over certain human abilities, and in doing so, disrupts existing responsibility practices, but also creates *novel* agential powers (affecting people in ways humans could not have done on their own), for which new responsibilities need to be *made*. For example, the existence of deepfakes presents an opportunity for impersonation and deception previously not possible by any human alone.

Our conceptual framing of the challenges to responsibility presented by AI allows us to make recommendations when it comes to the design of impact assessments. One such recommendation comes from our recognition that there may be some hidden tradeoffs that the traditional cost/benefit analysis in an impact assessment would not capture. In forthcoming work (Ganesh et al forthcoming) we discuss the knowledge/control tradeoff, namely, the consideration of whether the use of an AI-based system would likely generate enough additional knowledge to justify the loss of human control over a given outcome. For example, there has been a lot of money and effort poured into developing AI-based diagnostic tools for cancer detection, with the goal of potentially detecting smaller lesions. However, if one conducted the tradeoff analysis we are suggesting, then they might find that AI is better suited for research on diseases that currently lack clear diagnostic criteria. In these cases, the loss of control would actually be overcome by the potential gains in knowledge.

Additionally, in our work, we think critically about the concept of role responsibility, the idea in philosophical literature from Hart (1968) and others, that describes how human agents take on specific responsibilities as a function of their formal or social roles. Following this account, we propose that regulatory bodies also require that companies construct specific roles, i.e. roles tasked with ensuring fairness, technical accuracy, and safety, as well as community-based roles that collect and analyze reported harms, as well as surface potential new harms. These roles could report out to regulators on a regular basis. We argue that this forward-looking form of governance would be more nimble and effective than the risk-based strategy proposed by the EU AI Act, in which certain use cases are classified as high risk versus low risk, with bans or constraints of the particular use case prescribed on this basis. The risk-based classification approach ignores the potential for a type of AI-based technology's risk profile to *change over time* due to emergent agential powers. It also overstates the ability of regulators to be aware of every single use case, and anticipate harms well enough in advance to place a given AI-based system within the appropriate risk category.

A question not yet addressed in our work, and often missed in discussions of impact assessments or audits in general, is the question of *who* should be subject to audits. Given the issue of "many hands" in AI, it is unclear whether it should be the designers of the AI-based system, the deployers, or both. Design of any accountability scheme in this space needs to be careful to avoid the uneven distribution of legal liability because it will create market incentives that other companies will easily fill. For example, the emergence of both financial data brokers and nonbank "shadow banking" companies in the US can be attributed directly to the stringent regulations posed on banks after the financial crisis in 2008[26]. In addition to concerns about market incentives, any constructions of audits or impact assessments should be careful not to worsen existing power imbalances, i.e. making it easier for well-resourced companies to meet regulatory requirements.

In the era of steamboats, licensing requirements filled in some of the gaps created by "many hands". As mentioned earlier, there has been some discussion of requiring licensing of AI models, as well as software engineers/data scientists. However, the vast expansion of the many hands and many things problem, when compared to the steamboat era, suggests that there would still be many gaps remaining. This would be best illustrated with an example. Let's say we decide that any model that is designed needs to be licensed, and that the license should clearly state the intended use of the model. If one section of Google is developing the algorithms, and the other section of Google is deploying them for different use cases, there would be a conflict of interest, not too different from hedge funds, who both originated loans, and earned fees for fixing the loans when they were in default. Also, it might be counterproductive to add another hand to the many hands problem, in the form of a licensing board, which would potentially need oversight as well. These considerations suggest the need for innovative constructive governance methods in order to avoid potentially making AI-based systems even less safe.

[26] Illustrating this point, in 2020, the share of mortgage originations by nonbank lenders leapt to nearly 70 percent (Bhattacharya et al, 2021).

The administrative scope of the policy proposals we have highlighted may seem unwieldy and unrealistic, but governments have a vested interest in ensuring that they understand and evaluate the safety profiles and potential harms of these AI-based systems, since unsafe and poorly governed technologies often weaken adoption, undermine public trust, and amplify risk of catastrophic failure in critical systems. In discussing governance reponses to steamboat accidents, some have argued that it was the introduction of insurance that ultimately improved the safety of steamboats. In the case of AI governance, insurance could also be one way of improving outcomes, without placing too much pressure on already encumbered governments. While the viability of insurance should certainly be explored, constructive and innovative governance will be needed to make sure that the same issues of moral hazard and adverse selection that were present in the case of mortgage insurance, and ultimately led to the collapse of the global housing market, are not replicated.

There are a number of challenges that threaten the practical viability of the regulatory option. One such challenge is the slow pace of regulation. It took 30 years for steamboat regulation to be passed in the US. Given the fast pace of innovation, if it takes another 30 years to pass regulatory measures around AI, then the proposed requirements may be obsolete by then. The good news is that some of the fundamental debates of that time, such as the question of whether or not the government can regulate private business at all, have been settled. Even still, the passage of any legislation on this will require political will and prioritization over many other pressing issues. It remains to be seen if AI regulation springs to the top of such a priority list any time soon.

Additionally, the "inspectors" of audited systems would have to have a particular technical expertise to ensure that they are able to interrogate systems and report outcomes in the appropriate manner. This may be a challenge in government, however there are efforts in the US government aimed at combatting the lack of technical expertise, such as the creation of the US digital service, and recent technologist hiring programs at the FTC and CFPB. None of these obstacles are unprecedented, nor did these prevent the construction of effective governance mechanisms for steamboats and the many other new technologies that followed. We suggest they be seen as *tasks* for AI governance, not *roadblocks*.

Conclusion

In our paper, we turn to the history of technology to advocate that the challenges of governing AI are no more intractable than those posed by previous technologies. By highlighting the parallels between the issues of causal opacity and "many hands" presented by steamboats, and similar challenges posed by AI, we convey an optimism about the ability of our current challenges to be solved through similarly creative and constructive governance measures. We note the following lessons that can be taken from history. For one, the process of regulation is not linear and requires trial and error to accomplish its necessary aims. Many of those who remain skeptical about AI governance incorrectly portray regulation as being something that needs to be implemented at one snapshot in time. Instead, as technology and human activity with it co-evolve, the guardrails and policies needed to ensure their safety must continue to evolve as well. Second, we note that regulation is not the only effective means for ensuring that AI is safe and enhances rather than diminishes human flourishing. Through situating current governance proposals within Maust's taxonomy, we show the promise of the information and mechanical options currently being undertaken by the US, UK, and EU to combat causal opacity, while noting the importance of managing conflicts of interest that may arise due to the increasing industry capture of academia. Third, we argue that the steamboat case study highlights the need for independent government auditors in the case of AI, and indicates the potential usefulness of licensing mechanisms. However, we note that AI has made the issue of many hands vastly more challenging than in the era of steamboats, and that innovative approaches to governance will be needed to avoid the creation of perverse market incentives and exacerbation of existing power imbalances. Finally, as we move forward in developing global policies around AI, we suggest that the lack of global standardization of AI regulation be viewed as an opportunity rather than a hindrance for AI governance. During the steamboat era, Britain was inspired by US regulatory efforts to pass comprehensive steamboat regulation, while the US was inspired by Britain and France to adopt certain engineering standards and penalties for steamboat inspectors. In current discourse in AI, some suggest that there will be competing visions for AI governance, and we will have to see which vision 'wins out'. This sets up a false and ahistorical picture of technology governance as a zero-sum game. Instead, AI governance would be better off if we viewed competing global efforts as a path to learning and promoting innovation in governance itself.

Bibliography

Ada Lovelace Institute. 2022. "Algorithmic Impact Assessment: A Case Study in Healthcare." https://www.adalovelaceinstitute.org/report/algorithmic-impact-assessment-case-study-healthcare.

Armstrong, J., and D. M. Williams. 2003. "The Steamboat, Safety and the State: Government Reaction to New Technology in a Period of Laissez-Faire." *The Mariner's Mirror* 89 (2): 167–84. https://doi.org/10.1080/00253359.2003.10659284.

Barrett, Lisa Feldman, Ralph Adolphs, Stacy Marsella, Aleix M. Martinez, and Seth D. Pollak. 2019. "Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements:" *Psychological Science in the Public Interest*, July. https://doi.org/10.1177/1529100619832930.

Bender, Emily M. 2022. "Human-like Programs Abuse Our Empathy – Even Google Engineers Aren't Immune." *The Guardian*, June 14, 2022, sec. Opinion. https://www.theguardian.com/commentisfree/2022/jun/14/human-like-programs-abuse-our-empathy-even-google-engineers-arent-immune.

Bhattacharya, Krishna, Akshay Kapoor, and Ayush Madan. n.d. "Five Trends Reshaping the US Home Mortgage Industry | McKinsey." Accessed August 6, 2022. https://www.mckinsey.com/industries/private-equity-and-principal-investors/our-insights/five-trends-reshaping-the-us-home-mortgage-industry.

Birhane, Abeba, Pratyusha Kalluri, Dallas Card, William Agnew, Ravit Dotan, and Michelle Bao. 2022. "The Values Encoded in Machine Learning Research." arXiv. https://doi.org/10.48550/arXiv.2106.15590.

Brennan, Jenny. 2019. “Facial Recognition: Defining Terms to Clarify Challenges.” Accessed August 6, 2022. Ada Lovelace Institute (blog). https://www.adalovelaceinstitute.org/blog/facial-recognition-defining-terms-to-clarify-challenges/.

Brockmann, R. John. 2002. *Exploding Steamboats, Senate Debates, and Technical Reports: The Convergence of Technology, Politics, and Rhetoric in the Steamboat Bill Of 1838*. Amityville: Taylor & Francis Group. Accessed August 6, 2022. ProQuest Ebook Central.

Buchanan Thomas C. 1967. *Black Life on the Mississippi: Slaves, Free Blacks, and the Western Steamboat World*. Chapel Hill, University of North Carolina Press.

Burke, John G. 1966. “Bursting Boilers and the Federal Power.” *Technology and Culture* 7 (1): 1–23. https://doi.org/10.2307/3101598.

Burton, Vernon O., Troy Smith, and Simon Appleford. n.d. “Economic Development: The Golden Age of the Steamboat, 1851-1900 | Northern Illinois University Digital Library.” Accessed August 6, 2022. https://digital.lib.niu.edu/twain/steamboat.

Centre for Data Ethics and Innovation. 2019. “The Centre for Data Ethics and Innovation Calls for Evidence on Online Targeting and Bias in Algorithmic Decision Making.” Last modified https://www.gov.uk/government/publications/the-centre-for-data-ethics-and-innovation-calls-for-evidence-on-online-targeting-and-bias-in-algorithmic-decision-making.

Centre for Data Ethics and Innovation. 2021. “The Roadmap to an Effective AI Assurance Ecosystem.” Last modified December 8, 2021. https://www.gov.uk/government/publications/the-roadmap-to-an-effective-ai-assurance-ecosystem/the-roadmap-to-an-effective-ai-assurance-ecosystem.

Coeckelbergh, Mark. 2020. “Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability.” *Science and Engineering Ethics* 26 (4): 2051–68. https://doi.org/10.1007/s11948-019-00146-8.

Contractor, Danish, Daniel McDuff, Julia Katherine Haines, Jenny Lee, Christopher Hines, Brent Hecht, Nicholas Vincent, and Hanlin Li. 2022. “Behavioral Use Licensing for Responsible AI.” In *2022 ACM Conference on Fairness, Accountability, and Transparency*, 778–88. FAccT ’22. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3531146.3533143.

Costanza-Chock, Sasha, Inioluwa Deborah Raji, and Joy Buolamwini. 2022. “Who Audits the Auditors? Recommendations from a Field Scan of the Algorithmic Auditing Ecosystem.” In *2022 ACM Conference on Fairness, Accountability, and Transparency,* 1571–83. FAccT ’22. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3531146.3533213.

Denault, David John. 1993. “An Economic Analysis of Steam Boiler Explosions in the Nineteenth-Century United States.” PhD diss., University of Connecticut.

Feathers, Todd. 2021. "AI Can Guess Your Race Based On X-Rays, and Researchers Don't Know How." *Vice* (blog). August 23, 2021. https://www.vice.com/en/article/wx5ypb/ai-can-guess-your-race-based-on-x-rays-and-researchers-dont-know-how.

Federal Register. 2021. "Request for Information and Comment on Financial Institutions' Use of Artificial Intelligence, Including Machine Learning." Last modified March 31, 2021. https://www.federalregister.gov/documents/2021/03/31/2021-06607/request-for-information-and-comment-on-financial-institutions-use-of-artificial-intelligence.

Feenberg, Andrew. 2010. *Between Reason and Experience: Essays in Technology and Modernity*. Inside Technology. Cambridge, Mass: MIT Press.

Floridi, Luciano. 2021. "The European Legislation on AI: A Brief Analysis of Its Philosophical Approach." *Philosophy & Technology* 34 (2): 215–22. https://doi.org/10.1007/s13347-021-00460-9.

Gifford, Donald G. 2018. "Technological Triggers to Tort Revolutions: Steam Locomotives, Autonomous Vehicles, and Accident Compensation." *Journal of Tort Law* 11 (1): 71–143. https://doi.org/10.1515/jtl-2017-0029.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. "Datasheets for Datasets." *Communications of the ACM* 64 (12): 86–92. https://doi.org/10.1145/3458723.

Hart, H. L. A. 2008. "POSTSCRIPT: RESPONSIBILITY AND RETRIBUTION." In *Punishment and Responsibility*, 2nd ed. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199534777.003.0009.

Heinrichs, Jan-Hendrik. 2022. "Responsibility Assignment Won't Solve the Moral Issues of Artificial Intelligence." *AI and Ethics*, January. https://doi.org/10.1007/s43681-022-00133-z.

Hunter, Louis C. 1943. "The Invention of the Western Steamboat." *The Journal of Economic History* 3 (2). Cambridge University Press: 201–20. doi:10.1017/S002205070008356X.

Hunter, Louis C.1949. *Steamboats on the Western Rivers: An Economic and Technological History*. Cambridge: Harvard University Press.

Kapoor, Sayash, and Arvind Narayanan. 2022. "Leakage and the Reproducibility Crisis in ML-Based Science." https://doi.org/10.48550/ARXIV.2207.07048.

Köhler, Sebastian, Neil Roughley, and Hanno Sauer. 2017. "Technologically Blurred Accountability?: Technology, Responsibility Gaps and the Robustness of Our Everyday Conceptual Scheme." In *Moral Agency and the Politics of Responsibility*. Routledge.

Leveson, Nancy G. 1992. "High-Pressure Steam Engines and Computer Software." In *Proceedings of the 14th International Conference on Software Engineering*, 2–14. ICSE '92.

New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/143062.143076.

Lima, Gabriel, Nina Grgić-Hlača, Jin Keun Jeong, and Meeyoung Cha. 2022. “The Conflict Between Explainable and Accountable Decision-Making Algorithms.” *In 2022 ACM Conference on Fairness, Accountability, and Transparency*, 2103–13. https://doi.org/10.1145/3531146.3534628.

Lundberg, Scott M., and Su-In Lee. 2017. “A Unified Approach to Interpreting Model Predictions.” In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4768–77. NIPS’17. Red Hook, NY, USA: Curran Associates Inc.

Mashaw, Jerry L. 2008. “Administration and ‘The Democracy’: Administrative Law from Jackson to Lincoln, 1829-1861.” *The Yale Law Journal* 117 (8): 1568–1693. https://doi.org/10.2307/20454693.

Matthias, Andreas. 2004. “The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata.” *Ethics and Information Technology* 6 (3): 175–83. https://doi.org/10.1007/s10676-004-3422-1.

Maust, Peter. 2012. “Preventing ‘Those Terrible Disasters’: Steamboat Accidents And Congressional Policy, 1824-1860,” PhD diss., Cornell University.

McLachlan, Scott, Burkhard Schafer, Kudakwashe Dube, Evangelia Kyrimi, and Norman Fenton. 2022. “Tempting the Fate of the Furious: Cyber Security and Autonomous Cars.” *International Review of Law, Computers & Technology* 36 (2): 181–201. https://doi.org/10.1080/13600869.2022.2060466.

Means, Dennis R. 1987. “A Heavy Sea Running: The Formation of the U.S. Life-Saving Service, 1846-1878,” *Journal of National Archives*. 19 (4): 223-43.

Mitchell, Margaret, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. “Model Cards for Model Reporting.” In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–29. https://doi.org/10.1145/3287560.3287596.

Nyholm, Sven. 2018. “Attributing Agency to Automated Systems: Reflections on Human–Robot Collaborations and Responsibility-Loci.” *Science and Engineering Ethics* 24 (4): 1201–19. https://doi.org/10.1007/s11948-017-9943-x.

Ostmann, Florian, and Cosmina Dorobantu. 2021. “AI in financial services”. *The Alan Turing Institute*. https://doi.org/10.5281/zenodo.4916041

Raji, Inioluwa Deborah, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. 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, 33–44. Barcelona Spain: ACM. https://doi.org/10.1145/3351095.3372873.

Raji, Inioluwa Deborah, I. Elizabeth Kumar, Aaron Horowitz, and Andrew Selbst. 2022. "The Fallacy of AI Functionality." In *2022 ACM Conference on Fairness, Accountability, and Transparency*, 959–72. Seoul Republic of Korea: ACM. https://doi.org/10.1145/3531146.3533158.

Raji, Inioluwa Deborah, Peggy Xu, Colleen Honigsberg, and Daniel Ho. 2022. "Outsider Oversight: Designing a Third Party Audit Ecosystem for AI Governance." In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 557–71. AIES '22. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3514094.3534181.

Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. 2016. "'Why Should I Trust You?': Explaining the Predictions of Any Classifier." In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–44. KDD '16. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2939672.2939778.

Rice, Stephen P. 1963. *Minding the Machine: Languages of Class in Early Industrial America*. Berkeley: University of California Press. Accessed August 6, 2022. ProQuest Ebook Central.

Rudin, Cynthia. 2019. "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead." *Nature Machine Intelligence* 1 (5): 206–15. https://doi.org/10.1038/s42256-019-0048-x.

Santoni de Sio, Filippo, and Giulio Mecacci. 2021. "Four Responsibility Gaps with Artificial Intelligence: Why They Matter and How to Address Them." *Philosophy & Technology* 34 (4): 1057–84. https://doi.org/10.1007/s13347-021-00450-x.

Selbst, Andrew D. 2021. "An Institutional View Of Algorithmic Impact Assessments." SSRN Scholarly Paper 3867634. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=3867634.

Sharkey, Noel. 2010. "Saying `No!' to Lethal Autonomous Targeting." *Journal of Military Ethics* 9 (December): 369–83. https://doi.org/10.1080/15027570.2010.537903.

Soper, Spencer. 2021. "Amazon Sued Over Crashes by Drivers Rushing to Make Deliveries," *Bloomberg*, November 12, 2021. https://www.bloomberg.com/news/features/2021-11-12/amazon-com-algorithms-blamed-in-crash-that-paralyzed-aspiring-doctor.

Tigard, Daniel W. 2021. "There Is No Techno-Responsibility Gap." *Philosophy & Technology* 34 (3): 589–607. https://doi.org/10.1007/s13347-020-00414-7.

The White House. n.d. "Join the Effort to Create A Bill of Rights for an Automated Society." Accessed August 6, 2022. https://www.whitehouse.gov/ostp/news-updates/2021/11/10/join-the-effort-to-create-a-bill-of-rights-for-an-automated-society/.

U.S. Army Corps of Engineers. n.d. "A History of Steamboats." Accessed August 6, 2022. https://www.sam.usace.army.mil/Portals/46/docs/recreation/OP-CO/montgomery/pdfs/5thand6th/ahistoryofsteamboats.pdf

US EEOC. n.d. "The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence to Assess Job Applicants and Employees." Accessed August 6, 2022. https://www.eeoc.gov/laws/guidance/americans-disabilities-act-and-use-software-algorithms-and-artificial-intelligence.

Veale, Michael, and Frederik Zuiderveen Borgesius. 2021. "Demystifying the Draft EU Artificial Intelligence Act." *Computer Law Review International* 22 (4): 97–112. https://doi.org/10.9785/cri-2021-220402.

Villasenor, John. 2019. "Products Liability Law as a Way to Address AI Harms." *Brookings* (blog). October 31, 2019. https://www.brookings.edu/research/products-liability-law-as-a-way-to-address-ai-harms/.

Whittaker, Meredith. 2021. "The Steep Cost of Capture." *Interactions* 28 (6): 50–55. https://doi.org/10.1145/3488666.

Yew, Rui-Jie, and Dylan Hadfield-Menell. 2022. "A Penalty Default Approach to Preemptive Harm Disclosure and Mitigation for AI Systems." In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 823–30. Oxford United Kingdom: ACM. https://doi.org/10.1145/3514094.3534130.