information

MDPI

Article

Corporate Governance of Artificial Intelligence in the Public Interest

Peter Cihon [1], Jonas Schuett [2] and Seth D. Baum [3,*]

[1] Independent Consultant, San Francisco, CA 94107, USA; petercihon@gmail.com
[2] Legal Priorities Project, Cambridge, MA 02139, USA; jonas.schuett@legalpriorities.org
[3] Global Catastrophic Risk Institute, Washington, DC 20016, USA
* Correspondence: seth@gcrinstitute.org

Abstract: Corporations play a major role in artificial intelligence (AI) research, development, and deployment, with profound consequences for society. This paper surveys opportunities to improve how corporations govern their AI activities so as to better advance the public interest. The paper focuses on the roles of and opportunities for a wide range of actors inside the corporation—managers, workers, and investors—and outside the corporation—corporate partners and competitors, industry consortia, nonprofit organizations, the public, the media, and governments. Whereas prior work on multistakeholder AI governance has proposed dedicated institutions to bring together diverse actors and stakeholders, this paper explores the opportunities they have even in the absence of dedicated multistakeholder institutions. The paper illustrates these opportunities with many cases, including the participation of Google in the U.S. Department of Defense Project Maven; the publication of potentially harmful AI research by OpenAI, with input from the Partnership on AI; and the sale of facial recognition technology to law enforcement by corporations including Amazon, IBM, and Microsoft. These and other cases demonstrate the wide range of mechanisms to advance AI corporate governance in the public interest, especially when diverse actors work together.

Keywords: artificial intelligence; corporate governance; public interest; technology governance; multistakeholderism

Citation: Cihon, P.; Schuett, J.; Baum, S.D. Corporate Governance of Artificial Intelligence in the Public Interest. *Information* **2021**, *12*, 275. https://doi.org/10.3390/info12070275

Academic Editor: Luis Martínez López

Received: 30 April 2021
Accepted: 23 June 2021
Published: 5 July 2021

1. Introduction

The corporate governance of artificial intelligence (AI) can benefit from input and activity from a range of stakeholders, including those both within and outside of the corporation. Several recent initiatives call for multistakeholder governance institutions that bring diverse stakeholders together to inform AI governance. Examples include activities of the Global Partnership on AI [1], the European Commission's High-level Expert Group on AI [2], and research by Cath et al. [3]. To date, less attention has been paid to the important opportunities for different stakeholders to contribute to AI corporate governance in their own right—outside the context of dedicated multistakeholder institutions. Those opportunities are the focus of this paper.

The importance of AI corporate governance is clear. Corporations play a major—perhaps the primary—role in AI research, development, and deployment. Corporate-affiliated researchers published over 50% more AI research papers than academics in the United States in 2018 [4]. Corporate applications of AI touch on many important public issues, including social justice, economic vitality, and international security. Looking ahead, some have proposed that AI could displace large portions of the human labor pool, resulting in chronic unemployment for many people as well as massive profits for AI companies [5]. Corporations are also active in the research and development of artificial general intelligence, a technology that some believe could transform the world in ways that are either radically beneficial or catastrophic [6]. How AI is governed within corporations is therefore of profound societal importance.

To the best of our knowledge, this paper is the first survey of the corporate governance of AI. As reviewed below, prior publications have focused on specific aspects of the topic. This paper offers a broad introduction to the topic and resource for a wide range of scholarships and initiatives to improve AI corporate governance.

Although we aim for this paper to be a broad overview of opportunities in AI corporate governance, it does have some areas of focus. One is on select large corporations at the forefront of AI research and development, in particular Alphabet (the parent company of Google), Amazon, Facebook, and Microsoft. These corporations merit attention because they exercise significant influence on both technological developments and emerging regulatory methods. Additionally, within our discussion of government activities, there is a particular focus on the European Union, which has arguably the most mature regulatory landscape for AI to date. Finally, because this paper is focused on the practical mechanics of AI corporate governance, it mostly focuses on machine learning, the dominant AI paradigm today. These areas of focus are important in their own right; they also serve as examples to illustrate more general points about AI corporate governance that are applicable to other companies, political jurisdictions, and AI paradigms.

Three running examples illustrate how different actors can influence AI corporate governance. The first is Google's involvement in Project Maven, a U.S. Department of Defense project to classify the content of drone video. In 2018, Google management pulled Google out of Project Maven following media coverage and worker protests. The second example is on the open publication of potentially harmful AI research. In 2019, OpenAI announced its new strategy for publishing such research [7], sparking further debate by, among others, Partnership on AI [8]. The third example is on facial recognition for law enforcement. In 2020, a nexus of activity from nonprofits, the public, governments, and corporate management prompted several companies, including Amazon, IBM, and Microsoft, to stop providing facial recognition technology to law enforcement agencies. Although the paper also discusses other examples, these three run throughout the text and highlight the interconnected influence of different actors on AI corporate governance.

The paper is organized as follows. Section 2 presents the definitions of key terms. Section 3 reviews the extant literature. Section 4 assesses opportunities to improve AI corporate governance for a variety of actors: management, workers, investors, corporate partners and competitors, industry consortia, nonprofit organizations, the public, the media, and government. Section 5 concludes.

2. Definitions

Broadly speaking, *corporate governance* refers to the ways in which corporations are managed, operated, regulated, and financed. Important elements of corporate governance include the legal status of corporations in a political jurisdiction, the relationship between investors and executives, information flows within and outside of the corporation, and specific operational decisions made throughout the corporation [9]. Many people within and outside of a corporation can influence how the corporation is governed. For this reason, we take a broad view on the range of actors relevant for the corporate governance of AI.

Our specific focus in this paper is on how AI corporate governance can be improved so as to better advance the *public interest*. The public interest can be defined in many ways, such as in terms of costs and benefits, or voter preferences, or fundamental rights and duties. Exactly how the public interest is defined can be important for AI corporate governance, as is seen in a variety of controversies over AI applications. This paper does not take sides on the most appropriate conception of the public interest, with one exception, which is to reject the view that corporations' sole aim should be to maximize shareholder profits [10], and instead argue that corporations have obligations to a wider range of stakeholders. We recognize that this position is not universally held in the field of corporate governance; however, it does reflect support from many business leaders [11]. Broadly, our aim is to clarify the mechanisms through which corporate governance can be improved to better advance the public interest.

The concept of stakeholders is also central to this paper. *Stakeholders* have been defined as "any group or individual who can affect or is affected by the achievement of the organization's objectives" [12] (p. 46). Our focus is specifically on those who can affect how a corporation governs AI. Those who are affected by AI but cannot act to affect it, such as members of future generations, are outside the scope of this paper, except insofar as their interests are part of the overall public interest. It is therefore perhaps more precise to say that we focus on *actors*, i.e., those who can act to affect AI corporate governance. Likewise, our approach also parallels, but ultimately differs from, the phenomenon of *multistakeholderism*, which refers to governance activities conducted with participation from multiple types of stakeholders such as governments, corporations, academia, and nonprofits [13]. Multistakeholderism commonly manifests via dedicated institutions that invite participation from multiple types of stakeholders. Cath et al. call for new multistakeholder AI governance institutions [3]. Existing examples include PAI and the OECD Network of Experts on AI, which bring together people from government, industry, academia, and civil society to advance the understanding and practice of AI governance. These are important institutions, and they share this paper's interest in participation from a wide range of actors. This paper diverges from multistakeholderism by focusing on the full range of opportunities available to different actors and not just the opportunities afforded by dedicated multistakeholder institutions. The paper's approach is perhaps more similar to the concept of *stakeholder capitalism*, which calls for corporations to be attentive to stakeholder actions and responsive to stakeholder interests [14].

Artificial intelligence has been defined in many ways. One prominent definition states that AI is an artificial agent that can "achieve goals in a wide range of environments" [15]. However, current AI systems only perform well in certain settings, especially simpler environments for which there are ample data [16]. For this paper, it suffices to employ a social definition: AI is what people generally consider to be AI. This is a bit of a moving target: as the technology has progressed, people's minimum standards for what they consider AI have risen [17]. This paper focuses on the computer techniques currently considered to be AI, which, in practice, are largely machine learning, as well as more advanced forms of AI that may be developed in the future.

For AI corporate governance, it is also helpful to define AI activities in terms of the *AI system lifecycle*, i.e., the sequence of activities that take an AI system from its initial conception to its final use. Attention to the lifecycle can help identify and clarify opportunities to improve AI corporate governance. Different actors and activities will have varying influence over different phases of the AI system lifecycle within a corporation. This paper uses the AI system lifecycle to more precisely describe the influence of these actors and activities. In general, efforts to improve AI corporate governance must affect at least one phase of the lifecycle—otherwise, there is no effect on any actual AI systems.

This paper uses an AI system lifecycle framework developed by the OECD Expert Group on AI [18]. Figure 1 illustrates the four phases of the framework. Phase 1 concerns research and design of the AI system. Researchers identify a task for their system, choose a style of model, define performance measures, and select relevant data or other input. This phase includes data collection, cleaning, quality (including bias) checks, and documentation. Phase 2 tests the system to assess performance. This includes testing that covers regression (speed slowdowns), the comparison of previous model behavior to new behavior, and performance across many metrics, e.g., accuracy and calibration measures. Phase 3 puts the system into production. This may include launch testing for real-world use cases, checking compliance with relevant regulations, checking compatibility with legacy software, and assigning responsibilities for managing the AI system. Once the system is deployed, this phase also includes evaluating initial user experience. Phase 4 operates and monitors the AI system in deployment, assessing its outputs and impacts based on the designers' initial intentions and performance metrics as well as ethical considerations. Problems are identified and addressed by reverting back to other phases or eliminating the AI system.

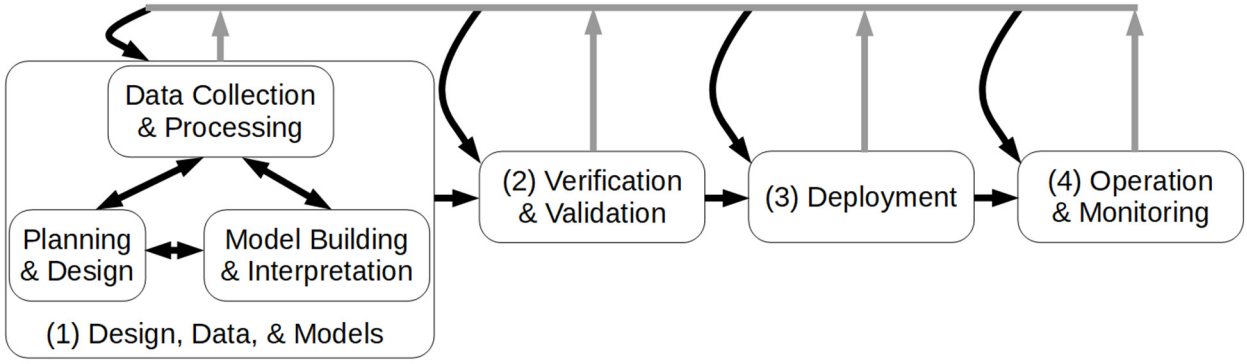

Figure 1. AI system lifecycle.

3. Prior Work

This paper sits at the intersection of literatures on corporate governance and AI governance. Corporate governance is a large field of scholarship with a long history. Good introductions are offered by Monks and Minow [19] and Gordon and Ringe [20]. AI governance is a smaller and relatively new field of study. For more work in this area, see, for example, works from the AI Now Institute [21], Data & Society [22], World Economic Forum [23], Future of Humanity Institute [24], as well as Calo [25].

One body of literature on AI corporate governance studies public policy proposals, primarily for new, dedicated governance bodies. Calo [26] calls for a federal body to address robotics policy. A similar proposal has been discussed in Europe by Floridi et al. [27]. The European Commission has recently proposed to establish a European Artificial Intelligence Board [28]. Scherer [29] outlines a proposal for a dedicated government agency and a voluntary certification scheme that incentivizes companies to submit to agency oversight in return for limited legal liability. Wallach and Marchant [30] propose a governance coordinating committee to support soft law governance that keeps pace with new and emerging AI. Erdélyi and Goldsmith [31] call for an international regulatory agency to address international AI challenges; Cihon et al. [32] argue that it is too soon to establish such an international structure and that further debate is first needed. Clark and Hadfield [33] propose a markets-based approach to AI safety regulation.

Some literature has analyzed existing regulations as they pertain to corporate AI. The E.U. General Data Protection Regulation (GDPR) has been of particular interest in this respect. For example, Wachter et al. [34] argue that the GDPR does not afford a "right to explanation" of automated decision-making, whereas Goodman and Flaxman [35] argue that it does. Another body of literature analyzes the European approach to AI regulation. Smuha [36] analyzes the emerging regulatory environment for making AI trustworthy. Thelisson et al. [37] and Stix [38] survey the broader regulatory landscape in the EU. There is also some literature on a number of national regulations. For example, Wagner et al. [39] analyze different corporate strategies for complying with algorithmic transparency requirements imposed by the German Network Enforcement Act.

There is also an extensive body of literature on specific policy instruments and governance approaches. For example, Senden [40] and Marsden [41] disentangle the concepts of soft law, self-regulation, and co-regulation. Kaminski [42] conceptualizes approaches between command and control regulation and self-regulation as "binary governance", while Pagallo [43] uses the framing of a "middle-out approach". Zeitlin [44] discusses the current state of transnational regulation within and beyond the E.U.

The legal liability of corporations for harms caused by AI systems has been another major focus. Broadly speaking, liability regimes aim to compensate victims of harms caused by products and, in turn, encourage producers to avoid the harms in the first place. AI liability regimes are generally not written specifically for the corporate sector, but in practice mainly affect commercial products. The exact form of liability regimes can vary substantially across jurisdictions and circumstances. Separate literatures discuss AI and

robotics under liability law in different jurisdictions, including the United States [29,45–48], the E.U. [49–53], and Germany [54–56]. Additionally, more theoretical approach considers whether liability regimes could handle extreme catastrophic risks from AI, such as those of potential long-term artificial general intelligence [57].

A variety of other AI corporate governance topics have also been studied. Buolamwini and Gebru [58] assess the efficacy of targeted audits and publicly shaming AI companies to address biases in facial recognition systems. More generally, Baum [59] explores the social psychology of AI developers as a factor in efforts to steer their work in pro-social directions. Belfield [60] details recent employee activism within the AI community and its impact on AI firms and technological development. Askell et al. [61] analyze competitive pressures on AI firms in terms of their societal impacts. Solaiman et al. [62] examine the societal implications of deciding to publicly disclose AI models, focusing on the case of OpenAI. Cihon [63] reviews the role of international technical standards in governing AI research and development. Baum [64,65] analyzes potential corporate efforts to manipulate public debate about AI risks. O'Keefe et al. [66] propose a novel method of corporate social responsibility that sees AI firms contribute to the public benefit. Avin et al. [67] and Ballard and Calo [68] use forecasting and roleplay methods to study potential future behaviors of actors affecting corporate AI.

A large number of articles published in magazines such as the *Harvard Business Review* and *MIT Technology Review* offer practical insights for managers on governing both AI development and adoption within firms. Hume and LaPlante [69] analyze how companies can manage biases and risks along the AI building process. Tiell [70] recommends corporations establish an ethics committee. Chamorro-Premuzic et al. [71] offer a step-by-step approach on how companies can build ethical AI for human resources. Fountaine et al. [72] outline how management should build AI-powered organizations. Abbasi et al. [73] analyze how companies can mitigate the risks of automated machine learning. Hao [74,75] urges AI companies to actually implement their ethical guidelines, while also emphasizing how difficult this will be.

Consulting firms have also published reports on AI corporate governance. Burkhardt et al. [76] of McKinsey describes how Chief Executive Officers (CEOs) can guide employees to build and use AI responsibly. Cheatham et al. [77], also of McKinsey, discuss how managers can mitigate AI risks. Ransbotham et al. [78] of the Boston Consulting Group survey more than 2500 corporate executives on AI topics including how companies are managing AI risks. Several major accounting firms have developed governance frameworks to promote ethical AI [79–81]. Deloitte [82] reports on AI risk management in the financial industry. Accenture [83] covers corporate AI ethics committees.

4. Actor-Specific Opportunities to Improve AI Corporate Governance

A variety of actors can improve AI corporate governance so as to better advance the public interest. This section considers nine types of actors. Three are internal to the corporation: managers, workers, and investors. Six are external: corporate partners and competitors, industry consortia, nonprofit organizations, the public, the media, and governments.

Although presented separately for clarity, these actors interact, overlap, and co-exist in practice. These various types of actors have important interactions. For example, the media can channel worker influence within firms, facilitate public pressure, and precipitate government action. Actors have the potential to overlap, for example, with governments publishing media reports or taking over management of a company. Ultimately, all actors co-exist within political cultures, which may vary by country and over time [84]: although the following sections describe actions that each actor could take to improve AI corporate governance, we do not offer analysis of the feasibility nor the desirability for such actions within their political cultural contexts.

4.1. Management

Management, as the term is used in this paper, includes all personnel with authority and oversight over other personnel, from top C-suite executives to mid- and lower-level managers. Management is an important—perhaps the most important—type of actor in corporate governance. Management establishes policies, implements processes, creates structures, and influences culture, all of which impact AI development.

One way management can advance AI in the public interest is by establishing corporate policies. One type of policy is strategic objectives. For example, management could establish the objectives of pursuing AI development where it is clearly in the public interest and avoiding contentious settings such as law enforcement. Another type of policy is ethics guidelines that specify how the corporation should develop and use AI and related technologies. Recently, management at many companies have established AI ethics guidelines, including Google, IBM, Microsoft, and OpenAI [85]. An ongoing challenge is to translate ethics guidelines into AI practice [86]. The translation process can include more operational policies on the details of how a company should develop and use specific AI techniques. Likewise, a concern is that published principles could create the appearance of AI corporations acting in the public interest without them actually doing so [87].

Management can also enact processes that translate policies into practice. These processes can take many forms. For example, management can establish new review processes for AI or augment existing review processes, such as those conducted by compliance and risk management teams. Additionally, management can encourage or require the use of documentation methods that generate and distribute information needed to ensure compliance with AI principles [88]. Notable examples include Datasheets for Datasets [89], a standardized reporting document for dataset features, and Model Cards [90], an approach to consistently describing an AI model and its intended use case. These processes could be improved if management were to review their efficacy and publicly share best practice.

Management activity on policies and processes is seen, for example, in the caution of OpenAI on publishing potentially harmful AI work. In 2019, OpenAI released their GPT-2 language model in phases out of concern about its potential harmful applications [7,62]. OpenAI created a review process to evaluate the social impacts of earlier releases before determining if and how to release more advanced versions of GPT-2. OpenAI's discussion of its phased release [7] references the OpenAI charter [91], a policy document that expresses the principle of factoring safety and security concerns into decisions of what work to publish. (Note: the charter was published in 2018, when OpenAI was a nonprofit.) Although authorship of the charter is attributed to "OpenAI", it is likely that OpenAI management played a central role in drafting and approving the document, which anchors the organization's "primary fiduciary obligation" [92]. Additionally, the OpenAI GPT-2 team includes a mix of workers and management, including OpenAI co-founder and Chief Scientist Ilya Sutskever; thus, it can be inferred that management was likely involved in the review process. In summary, the GPT-2 release appears to demonstrate how management may translate policies into processes to support AI development and use in the public interest.

Management can also create structures within the company dedicated to advancing AI in the public interest. Such structures can perform oversight, make recommendations, and provide expertise to people throughout the company. They can consist of company staff and often interact with numerous teams across the organization. Prominent examples include the Microsoft advisory committee AI, Ethics, and Effects in Engineering and Research, the compliance-oriented Microsoft Office of Responsible AI, the Google Responsible Innovation Team, the AI Principles working group within the Google Cloud division, and a Facebook team of policy managers that work with product teams on fairness and explainability problems. Alternatively, the groups can consist of external advisors, such as the Google DeepMind Ethics & Society division's group of external advisors, the Axon AI and Policing Technologies Ethics Board, and the short-lived Advanced Technology External Advisory Council at Google.

Thus far, these dedicated structures have had mixed success. One success came at Axon. Its ethics board advised against the company selling facial recognition to law enforcement; management followed this advice [93]. A failure occurred at Google, which disbanded its Advanced Technology External Advisory Council soon after its launch amid outcry about its membership [94]. Overall, these structures are relatively new and not yet in wide use, and much is still being learned about them. Nonetheless, one early lesson is that AI governance teams ought to be interdisciplinary. Regardless of where such a team may be within the reporting structure, it may be expected to include lawyers, ethicists, data scientists, engineers, program managers, and other diverse occupational perspectives.

Management can also build AI governance functions into pre-existing structures. Important areas for this may be in compliance and risk management. Current compliance and risk management teams may focus less on AI and more on established risks such as computer security [95]. However, as governments increase their policy attention to AI, the need for corresponding activity within corporations will increase. It is likely that the pre-existing corporate structures could build expertise in AI risks over time, as the field of AI corporate governance matures, as standards are published, and as regulations enter into force. Forward-thinking management can advance this process by setting the groundwork, such as by building AI expertise into pre-existing structures.

Finally, management can help cultivate a corporate culture that supports AI development in the public interest. Corporate culture can play a major role in how a company develops and uses AI [59,96]. Employee onboarding and training could include a focus on responsible AI development [97]. Recruiting efforts could select for, or aim to instill, knowledge of responsible AI development methods. Employee performance reviews and metrics could incentivize these methods' use, from concretely assessing bias in training data at the design phase to more broadly upholding a culture of responsible development. On the latter, OpenAI has tied compensation levels to adherence to its charter [98]. However, it is unclear what additional steps dominant AI firms are now taking to instill their AI principles into corporate culture.

4.2. Workers

We use the term workers to refer specifically to people who work at the company and do not have managerial authority. This includes, in their subordinate relationship to top management, lower- and mid-level managers. Workers include employees and contractors, both of which are common at AI companies. A wide range of workers affect AI, including researchers, engineers, and product developers. Despite being expected to follow directions from management, workers at AI firms have considerable power to shape corporate governance. Workers are often left with significant latitude to determine corporate activity within their areas of focus, and management is often (although certainly not always) influenced by worker suggestions.

Workers can influence AI corporate governance both directly, through their actions affecting AI systems, and indirectly, by influencing management. While management makes many governance decisions, especially high-level decisions for the corporation and its divisions, many other decisions are left to workers, especially on the specifics of AI design and implementation. Worker opportunities for direct influence may be especially robust at earlier stages of the AI system lifecycle and at corporations and divisions whose management offer workers wide latitude for decision-making. Likewise, worker opportunities for indirect influence may be greatest at corporations and divisions whose management is especially receptive to worker input.

Some of the best opportunities for worker direct influence may be for workers in groups conducting fundamental research, such as Facebook AI Research, Google Brain and DeepMind, Microsoft Research, and OpenAI. Workers in these groups may have significant autonomy from management to pursue their work as they see fit. Indeed, these workers may be influenced less by management and more by academic norms, research fashions, reputational concerns, conference requirements, journal expectations, and their

own personal values. Likewise, those seeking to influence corporate AI researchers may find good opportunities via the broader field of AI, such as at leading conferences. For example, as of 2020, the NeurIPS conference uses an ethics review process and requires papers to include a social impact statement [99]. These activities can be important for AI corporate governance due to the significant autonomy of corporate AI researchers.

Workers also have significant opportunities to indirectly affect AI corporate governance by influencing management. However, these opportunities can be risky for workers because of managements' control over—or at least considerable influence on—workers' employment and advancement within the corporation. In general, activities that involve a higher degree of worker commitment and risk of reprisal by management will tend to have a greater effect on corporate governance. Low-commitment, low-risk activities can be as simple as raising concerns in project meetings over issues of ethical AI development. These activities tend to be low-profile and not well-documented; colleagues at AI companies inform us that these activities are nonetheless common. More ambitious and risky activities tend to be less common but more visible and more well-documented. These activities can include circulating letters critiquing corporate activity and calling for change, whistleblowing, organizing walkouts, forming unions, and more [60].

Likewise, the extent of indirect worker influence is shaped by management's receptiveness to worker input and on related management decisions regarding corporate policy and culture. In extreme cases, management can fire workers who push back against management AI corporate governance decisions. For example, Google fired its Ethical AI team co-lead, Timnit Gebru, following a disagreement over the publication of a paper critical of the company's research on large AI language models [100]. Additionally, several Google employees claim to have been fired as retribution for labor organizing, in possible violation of U.S. labor law [101]. Subtler dynamics include such matters as whether workers have dedicated spaces to organize and articulate their views. For example, Google has internal discussion forums for workers, although management recently hired a team to moderate them [102]. Google management also recently eliminated its regular meetings where employees could address executives [103]. In general, workers will have greater indirect influence on AI corporate governance when they can organize and express views to management without fear of retaliation.

The size of the labor pool also affects both direct and indirect worker influence. Some governance goals may benefit from a large labor pool, such as the goal of solving difficult technical problems in orienting AI toward the public interest. Greater availability of worker talent may make these problems easier to solve. On the other hand, a larger labor pool can make it difficult for workers to self-organize and reach consensus. Likewise, a large labor pool relative to demand for their labor reduces indirect worker influence on AI systems via their influence on management [60].

At present, there is a shortage of talent in the computer science and engineering dimensions of AI, giving workers in these areas considerable indirect influence. These workers are hard to hire and to replace upon firing; therefore, management may be more inclined to accept their demands. This influence could fade if the labor market changes due to increased university enrollment in AI courses and the many government calls for training more people in AI [104] (pp. 111–126); [105]. Labor demand could also shrink if the applications of AI plateau, such as due to a failure to overcome limitations of current deep learning algorithms [16] or due to the rejection of AI applications on moral, legal, or social grounds. For now, though, demand for AI is robust and growing, giving AI scientists and engineers substantial power.

One powerful pattern of indirect worker influence starts with whistleblowing and continues with widely signed open letters. Workers with access to information about controversial AI projects leak this information to media outlets. Subsequent media reports spark dialogue and raise awareness. The media reports may also make it easier for other workers to speak publicly on the matter, because the workers would no longer have to shoulder the burden of being the one to make the story public. The open letters then provide

a mechanism to channel mass worker concern into specific corporate governance actions to be taken by management. (See also Sections 4.1 and 4.8 on the roles of management and the media.)

This pattern can be seen in several recent episodes at Google. In 2018, Google's participation in Project Maven, a U.S. Department of Defense project to use AI to classify the content of drone videos, was anonymously leaked to *Gizmodo* [106]. The *Gizmodo* report does not explicitly identify its sources as Google workers, but this is a likely explanation. Subsequently, over 3000 employees signed an open letter opposing Google's work on Project Maven [107]. Google management later announced it would leave Project Maven and publish principles to guide its future work on defense and intelligence projects [108]. Additionally, in 2018, *The Intercept* reported Google's work on Project Dragonfly, a Chinese search engine with built-in censorship [109]. *The Intercept* report was also based on an anonymous source that appears to be a Google worker. Subsequently, over 1000 employees signed a letter opposing the project [110]. Google management later ended the project [111].

A somewhat similar pattern is observed in a 2018 episode involving sexual harassment at Google. A *New York Times* investigation of corporate and court documents and interviews of relevant people found that Google had made large payments to senior executives who left the company after being credibly accused of sexual harassment [112]. Soon after, Google workers organized walkouts in which thousands of workers walked out in support of corporate policy changes on harassment and diversity [113]. The organizers referenced the *New York Times* report but did not specify the extent to which the walkout was motivated by the report. The organizers later wrote that Google management made some but not all of their requested policy changes [114].

These sorts of worker initiatives are not always successful. In 2018, an unspecified number of Microsoft employees published an open letter calling on Microsoft to abandon its bid for the Joint Enterprise Defense Infrastructure contract, a U.S. Department of Defense cloud computing initiative [115]. Microsoft did not abandon its bid, although Microsoft President Brad Smith did respond by articulating Microsoft policy on military contracts [116]. Additionally, in 2018, hundreds of Amazon employees signed a letter demanding the company stop selling facial recognition services to law enforcement [117]. Management did not stop. Again in 2018, approximately 6000 Amazon employees signed a letter calling on the company to stop using AI for oil extraction. The letter was accompanied by a shareholder resolution making the same argument—an example of investor activity (Section 4.3). Again, management did not stop [118].

4.3. Investors

Corporations take investments in a variety of forms, including by selling shares of stock or issuing bonds. Investors are important because AI is often capital-intensive, requiring extensive funding for research, development, and deployment. Shareholders are the investors with the most capacity to influence corporate governance and are therefore the focus of this section. Indeed, a central theme in corporate governance is the principal–agent problem in which the principals (i.e., shareholders) seek to ensure that their agents (i.e., corporate management) act in the principals' best interests rather than in those of the agents. In contrast, issuers of bonds are generally less influential, in part because markets for bonds are highly competitive—a corporation can readily turn to other issuers instead of following one issuer's governance requests.

Investors can influence corporations in several ways. First, investors can voice their concerns to corporate management, including at the annual shareholder meetings required of U.S. companies. Investor concerns can, in turn, factor into management decisions. Second, shareholders can vote in shareholder resolutions, which offer guidance that is generally non-binding but often followed [19] (p. 117). Indeed, even resolutions that fail to pass can still succeed at improving corporate governance; evidence for this has been documented in the context of environmental, social, and governance (ESG) issues [119]. Third, shareholders can replace a corporation's board of directors, which has ultimate

responsibility to manage the corporation, determines strategic direction, and appoints the CEO. For example, shareholders could seek to add more diversity to a board, noting that boards with increased diversity are associated with greater support for corporate social responsibility efforts [120]. Fourth, investors can signal disapproval with corporate governance practices by selling off their investments, and, perhaps, investing in a better governed competitor. Fifth, shareholders can file lawsuits against the corporation for failing to meet certain obligations [121]. These lawsuits are often settled in ways that improve corporate governance [19] (p. 117).

In principle, investors can wield extensive power over a corporation via their control over the board of directors. If management does not follow investor concerns or shareholder resolutions, the shareholders can replace the board with people who will. In practice, however, investor power is often limited. Efforts to replace a board of directors are expensive and rare [19] (p. 117). One study found that activist investors launched 205 campaigns in 2019 and won only 76 board seats [122]. This reality gives management substantial latitude in corporate governance. Nonetheless, management often does respond to investor preferences, especially, but not exclusively, when their preferences affect the company's stock price.

The power of investors is also influenced by the availability of alternative investment options. A diverse market of AI investment opportunities would provide investors with opportunities to shift their assets to companies that further the public interest. Current market prospects are mixed. On one hand, much of the sector is dominated by a few large companies, especially Google (Alphabet), Amazon, Facebook, and Microsoft. On the other hand, there is also a booming AI start-up scene today; one study identified 4403 AI-related companies that received a total of USD 55.7 billion in funding in the year ending July 2019 [4] (p. 91). Companies such as H20 and Fiddler specifically aim to advance explainable AI systems, creating additional opportunities for investors to promote AI in the public interest.

Investor initiatives should be well-informed by the state of affairs in the company. This requires some corporate transparency. For example, the U.S. Securities and Exchange Commission (SEC) requires companies to disclose investor risk factors [123]. In its disclosure, Google (Alphabet) lists AI-related "ethical, technological, legal, regulatory, and other challenges." Amazon cites uncertainty about the potential government regulation of AI. Facebook mentions that AI creates a risk of inaccuracies in its community metrics. Microsoft lists AI as a risk of "reputational harm or liability". However, at each company, AI is only a small portion of the overall disclosure, suggesting that the companies see AI as a minor area of risk. Investors could consider the possibility that the companies are not giving AI risks the attention they deserve.

The efficacy of investor initiatives as an approach to improving AI corporate governance depends on the willingness of investors to take the matter on and the investors' degree of influence within the company. Investors are diverse and have many interests. Investors with an existing interest in ESG may be especially receptive to promoting AI in the public interest. For example, Hermes, an ESG-oriented investment management business, has written on responsible AI [124] and participated in an investor initiative to create a Societal Risk Oversight Committee of the Board at Alphabet [125]. That initiative ultimately failed [126], in part because it lacked the support of Alphabet's founders. Alphabet is structured such that their founders retain a majority of shareholder voting power even though they do not own a majority of the shares; Facebook is structured similarly [127]. Although Alphabet and Facebook are extreme cases, in general, investor initiatives will tend to be more successful when they are supported by investors who own a larger portion of shareholder voting power. This applies to public corporations that have issued stock. Investors in private corporations may be especially influential, especially for smaller firms, which often have less access to capital markets. Venture capital firms seeking to promote the public interest may be especially successful in improving AI corporate governance among smaller firms.

The limited influence of shareholder resolutions can also be illustrated by the failed attempt to restrict Amazon's sale of facial recognition technology to the government. In 2019, shareholders expressed their concern that Rekognition, Amazon's facial recognition service, poses risk to civil and human rights, as well as shareholder value. They requested the Board of Directors to prohibit sales of such technology to government agencies [128] (pp. 18–19). In 2020, another resolution requested an independent study of Rekognition, including information about the extent to which such technology may endanger civil rights and is sold to authoritarian or repressive governments [129] (pp. 25–26). Both resolutions failed. Even though they would have been non-binding, Amazon tried to block the vote. This unusual attempt was ultimately stopped by the SEC [130]. Although these resolutions did not succeed in achieving reform, they demonstrate that shareholder activism has begun to focus on AI risks in particular.

In short, shareholders wield some influence in the corporate governance of AI. This influence is limited by the sheer volume and variety of risks that weigh on them: AI is not a top of mind. The most used activity available to shareholders thus far has been the resolution. Given that shareholder resolutions are difficult to pass and non-binding when passed, it is unclear if such activities will do much to change corporate governance aside from publicizing particular AI-related governance problems. Over time, as companies continue to emerge that seek competitive differentiation through responsible AI development and as shareholders, particularly institutional investors, continue to value ESG criteria and apply them to AI, the role of investors in responsible AI governance may continue to grow.

4.4. Corporate Partners and Competitors

Other corporations exert influence on a corporation developing or using AI in important ways. These other corporations can be direct competitors, themselves developing or deploying AI systems. Alternatively, they can be corporate partners (or, for brevity, "partners") that have contractual relationships with said company. Partners can be, among other things, suppliers, customers, or insurers. Partners can use their relationship with the AI company to influence it to advance the public interest.

Competing AI corporations can influence each other through direct market competition and in other ways. As classic economic theory explains, competition can result in greater market share for corporations whose products better advance the public interest. There are exceptions, including where there are negative externalities, i.e., harms of market activity that are not captured by market prices, and monopolies, i.e., where a large market share can be used to exclude competition and set relatively high prices. For example, direct competition to develop more powerful machine learning systems can result in better performance for important applications such as healthcare and transportation, but it can also result in more energy consumption and the externalities of global warming via the use of large amounts of computer hardware [131].

Competitors can also influence each other as peers in the overall field of AI [132]. One AI corporation's initiatives in the public interest can be adopted or adapted by other AI corporations. For example, in 2020, IBM announced that it would no longer sell facial recognition technology to law enforcement agencies [133]. Amazon [134] and then Microsoft [135] did the same shortly after. These announcements came amidst heightened attention to police misconduct sparked by the killing of George Floyd by the Minneapolis Police Department and subsequent widespread Black Lives Matter protests; therefore, it is possible that each company would have changed its behavior on its own without the others doing the same. However, in an interview, Microsoft's President explicitly recognized IBM's and Amazon's steps [135]. To some extent, the companies may have been jockeying for market share in the face of shifting public opinion, but they may also have been motivated by each other's example to advance the public interest.

Partners' ability to influence AI companies can depend significantly on their relative market power. The AI sector is led by some of the largest companies in the world. These companies are often in position to dictate the terms of their partner relationships; they have

sometimes used this power to ensure that AI is used in the public interest. For example, Google has limited the use of its facial recognition services to a narrow customer base via its Celebrity Recognition API, and has an extended terms of service to regulate how the technology is used [136]. Similarly, Microsoft vets customers for its facial recognition services; before its blanket 2020 policy was implemented, it reviewed and denied a California law enforcement agency's request to install the technology on body cameras and vehicle cameras [137].

Partners can impact the reputation of AI companies and, in turn, influence actions to protect that reputation. For example, Article One Advisors, a consulting firm, worked with Microsoft to develop its human rights policies through external engagement, and then publicized this work [138]. The publicity likely boosts Microsoft's reputation and incentivizes Microsoft to follow through on its public commitments. Corporate partners can also harm AI corporations' reputations. For example, Google attracted widespread criticism when Randstad, a staffing contractor, allegedly collected facial scans of homeless African Americans in order to improve the performance of Google's Pixel phone facial recognition features [139]. Reputation is important for large, public-facing corporations such as Microsoft and Google, making this a valuable tool for their corporate partners.

Insurers have distinctive opportunities to influence AI companies. When AI is not in the public interest, that can create substantial risks that may require insurance payouts. Insurers therefore have both the vested interest and the contractual means to compel AI companies to act in the public interest. For comparison, insurers have mandated the adoption of cybersecurity governance and risk frameworks [140,141]; they could do the same for AI. Doing so would improve corporate governance in the insured AI companies. Additionally, it could have further benefits by popularizing innovative practices for AI governance and risk management that are adopted by even uninsured companies. However, some AI risks are not readily handled by insurers, such as emerging risks that are difficult to quantify and price and risks that are too extreme, such as risks from long-term artificial general intelligence.

Finally, it should be noted that corporate partners and competitors consist of management, workers, and investors, whose influence parallels that of their counterparts in AI corporations as discussed in Sections 4.1–4.3. Workers, managers, and investors who seek to improve AI corporate governance may find additional opportunities in corporate partners and competitors. As an illustrative example, in 2020, FedEx investors pushed FedEx to call for the Washington Redskins American football team to change their name, given its racist connotations. FedEx is a major partner of the team. The initiative was successful: the team name will be changed [142]. This example is not from the AI industry, but it nonetheless speaks to the capacity for actors in corporate partners and competitors to affect positive change.

4.5. Industry Consortia

Industry consortia, as the term is used here, are entities in which multiple corporations come together for collective efforts related to AI governance. We define industry consortia to include entities that include more than just corporations. For example, PAI membership includes corporations, nonprofits, media outlets, and governmental bodies [143]. PAI is perhaps most precisely described as a multistakeholder organization, but it is also an industry consortium. The same holds true for other entities, such as the IEEE Standards Association, whose members include corporations and individuals [144].

Industry consortia can be instrumental in identifying and promoting best practices for AI in the public interest. AI corporations face many of the same challenges and issues. They likewise benefit from best practices being developed for the whole sector and then distributed to each corporation, instead of each corporation "reinventing the wheel". Industry consortia are well-positioned to serve as the entity that develops best practices for the whole sector. They can query member corporations on what has worked well—or poorly—for them, pooling their collective experience together. They can also conduct

in-house research on best practices, with researchers hired by the pooled funds of their member corporations. It may not be worthwhile for every AI corporation to hire their own in-house experts on various facets of AI in the public interest, but it may be worthwhile for the sector as the whole to do so. Industry consortia enable that to happen.

An illustration of these dynamics is seen in PAI's recent work on best practices on the publishing of potentially harmful AI research. PAI's work on this was prompted by the work of one of its members. Specifically, OpenAI released its language model GPT-2 in phases out of concern about its potentially harmful uses [62]. Soon after, PAI hosted discussions with OpenAI and other organizations about best practices in publishing potentially harmful research [145], launched a project to develop guidance [8], and advised Salesforce on the release of their language model CTRL [146,147]. PAI's status as an industry consortium has enabled it to advance publishing practices across organizations.

As best practices are formulated, industry consortia can take the additional step of formalizing them as standards. For example, the Consumer Technology Association (CTA) convened over 50 companies, some but not all of which were CTA members, to develop a standard for the use of AI in healthcare [148]. The IEEE Standards Association is also active on AI standards, as is the International Organization for Standardization [63], although the latter is not an industry consortium. By formalizing best practices into standards, industry consortia can enable corporations across the sector to improve their practices.

Best practices and standards developed by industry consortia can go on to play a legal or quasi-legal role. Governments sometimes enact policies requiring corporations to adhere to certain broad principles of conduct without specifying the details of what particular conduct does or does not meet these principles [149]. The best practices and standards formulated by industry consortia can fill in the details of good conduct. Additionally, regulatory agencies and courts handling liability cases sometimes treat compliance with industry best practices and standards as satisfactory, such that corporations meeting these practices or standards avoid regulatory fines or court judgments of liability to which they would otherwise be subject [150] (p. 17). This can dilute the public benefits of government action, but it also incentives corporations to meet or exceed these standards and practices, potentially bringing net gains for the public interest.

The above are examples of soft law, which can be defined as obligations that, although not legally binding themselves, are created with the expectation that they will be given some indirect legal effect through related binding obligations under either international or domestic law [151]. Soft law has been advocated for AI corporate governance due to its flexibility and ease of adoption [152,153]. In general, it is difficult for governments to create detailed and rigorous laws for complex issues such as those pertaining to AI. The dynamism of emerging technologies such as AI is especially challenging for the development and enactment of "hard" laws. Industry consortia are often better positioned than governments to master the details of the technology and its changes over time, due to the availability of expertise among consortium members. Furthermore, any more binding "hard law" measures enacted by governments are likely to draw on the particulars of pre-existing soft law instruments. These are additional reasons for industry consortia to pursue robust best practices and standards for AI corporate governance in the public interest.

Industry consortium activities do not necessarily advance the public interest. For example, they can pool the resources of member corporations to lobby governments for public policies and conduct information and public relations campaigns that advance industry interests at the public's expense. In other sectors, such lobbying has often been a major impediment to good public policy [154]. In the coming years, industry consortia could possibly present similar challenges to the public interest.

4.6. Nonprofit Organizations

Nonprofit organizations play several important roles in advancing AI corporate governance in the public interest, including research, advocacy, organizing coalitions, and education. Nonprofits can be advocacy organizations, labor unions, think tanks, political

campaigns, professional societies, universities, and more. The distinction between these types of organizations is often blurry, with one organization playing multiple roles.

To date, research has been a primary focus of nonprofit organizations working on AI corporate governance. Research contributions from nonprofit universities and think tanks are too numerous to compile here; many are in Section 3. What follows are some select examples of nonprofit research aimed at influencing corporate governance. Note that all universities mentioned in this section are nonprofit. Upturn, a nonprofit dedicated to advancing equity and justice in technology, worked with researchers at Northeastern University and University of Southern California to produce evidence of previously suspected illegal discrimination in housing advertisements served by Facebook [155]. Ranking Digital Rights (e.g., [156]) reports technology companies' human rights records and encourages companies to improve their performance. The Electronic Frontier Foundation reports companies' cooperation with government demands for censorship and also encourages them to improve their performance [157]. The AI Now Institute at New York University publishes reports to provide AI developers with suggestions to reduce bias and increase the public accountability of AI systems [158]. Finally, this paper is another work of nonprofit research on AI corporate governance.

Nonprofit advocacy efforts can often draw on such research. For example, a 2018 advocacy campaign by the nonprofit American Civil Liberties Union (ACLU) opposed Amazon selling facial recognition software to governments [159]. The campaign was supported by research on biases in the software by the ACLU [160] and prior research from a pair of researchers at Microsoft and the Massachusetts Institute of Technology [58]. The ACLU later reported that their campaign was unsuccessful [161]. However, in 2020, following anti-police brutality protests, Amazon stopped selling facial recognition software to law enforcement agencies, as discussed in Sections 4.4 and 4.7. The ACLU campaign and the research on which it drew may have laid the groundwork for Amazon's action.

Finally, nonprofits have conducted some work to build the field of AI in directions beneficial to public interest. Black in AI and AI4All are nonprofit organizations that promote diversity within the field of AI. Increased diversity could, in turn, help reduce bias in the design, interpretation, and implementation of AI systems. Additionally, the Future of Life Institute has hosted conferences on beneficial AI and built coalitions in support of open letters calling for AI in the public interest. These field-building initiatives are not specifically focused on corporate governance, but they have included people from AI corporations and are likely to have at least some effect on AI corporate governance.

One challenge facing nonprofit organizations is funding. This is a challenge for nonprofits working on all cause areas, and AI corporate governance is no exception. Firstly, nonprofits may struggle to raise the funds they need to advance their missions. Secondly, some AI nonprofits may turn to AI companies for funding, creating potential conflicts of interest [162,163]. Thirdly, where companies disagree with the nonprofits' aims, the companies can use their wealth to push back. Although such a dynamic has perhaps not been seen much to date in AI, it has been observed in other sectors, such as the tobacco industry pushing back on the link between cigarettes and cancer and the fossil fuel industry pushing back on the risks of global warming [64]. The extreme wealth of some AI corporations makes the potential for conflict of interest and the imbalance in resources particularly acute. Where these issues arise, nonprofits may fail to advance the public interest.

With that in mind, it can be helpful to distinguish between adversarial and cooperative nonprofit activity. Adversarial activity pushes AI companies in ways that the companies do not want. Cooperative activity proceeds in ways that the companies are broadly supportive of. Cooperative activity may tend to have a more limited scope, limited by the bounds of what companies are willing to support. On the other hand, adversarial activity may struggle to effect change if the companies do not agree with the proposed changes, and in some cases could backfire by galvanizing opposition. Whether adversarial or cooperative approaches are warranted should be assessed on a case-by-case basis.

4.7. The Public

The public, as the term is used here, refers to people acting in ways that are non-exclusive in the sense that broad populations can participate. In the context of AI corporate governance, the primary roles of the public are (1) as users of AI technology, including when the users pay for it and when it is free to them, subsidized by advertising, and (2) citizens who can vote and speak out on matters of public policy. Although not discussed in this section, members of the public can also impact AI corporate governance indirectly by exerting influence on other members of the public.

Users of AI technology can improve AI corporate governance by choosing goods and services that are in the public interest. In other (non-AI) industries, customers are often—although not always—willing to pay more for branded ethical standards [164]. AI users may also be willing to pay more; or, when they are using the technology for free, they could accept a product that has higher ethical standards but is inferior in other respects. This effect can even determine which technologies become dominant, rejecting certain uses and prioritizing others—regardless of how they are marketed [165]. One example in this direction is the gradual shift in social media platform popularity from public platforms such as MySpace, Facebook, and Twitter toward private messaging platforms, such as Snapchat and WhatsApp [166].

Citizens can improve AI corporate governance by supporting good (and opposing bad) AI public policies. They can do this by speaking up, such as in anti-racism and police brutality protests that have influenced AI facial recognition practices, and by voting for politicians who will enact good policies. Public opinion has played an important role in regulatory responses to other advanced technologies such as genetically modified organisms [167,168]. Growing public concern about digital technology, dubbed "techlash", has prompted calls for antitrust and other policy response. Thus far, AI has not been extensively regulated, although further shifts in public opinion could help to change this.

Public opinion has played an important role with regard to the sale of facial recognition software to law enforcement. As discussed in Sections 4.4 and 4.6, following 2020 protests against racism and police brutality, several AI companies moved away from providing facial recognition tools for law enforcement. It is worth noting that the protests garnered broad public support [169]. Therefore, the AI corporations' responses show how public protests and changes in public opinion can advance AI corporate governance in the public interest.

Finally, members of the public can voice their views about AI, prompting changes. For example, the initial launch of Google Glass in 2014 was widely criticized for violations of privacy [170]; it was discontinued and subsequently relaunched for industrial and professional users instead of for the general public [171]. Google Photos, an AI photo classification system, sparked public outcry for labeling a person with dark skin as a gorilla, prompting Google Photos to remove gorilla and other non-human primate terms from its lexicon [172]. In general, public pressure will tend to be more pronounced for highly visible brands [173,174]; the same is likely to also apply for AI companies.

The public faces several challenges to supporting the corporate governance of AI in the public benefit. One is the complexity of AI issues, which makes it hard for people lacking specialized training to know what the issues are or what stances to take. This can be mitigated by efforts for public education, such as Elements of AI [175], an online course which aims at educating 1% of European citizens in the basics of AI. Despite such initiatives, public education remains a difficult challenge due to the complexity of the issues and the competition for public attention. Likewise, public education can take time, in which case it may be most valuable in the medium term. (On medium-term AI issues, see Ref. [176])

Furthermore, some AI issues are important but arcane and not conducive to media coverage or other means of capturing public attention. This holds in particular for low-visibility AI companies, including those that do not market to the public but instead sell their AI to governments or other companies.

In some cases, AI technology users may be relatively disinclined to opt for the more ethical option due to the difficulty of switching from one AI product to another. Impediments can include (1) significant learning curves, as is common for software in general, (2) transition costs, such as the need to re-upload one's photos and other information to a new site, or the need to inform one's contacts of their new email address, and (3) network effects, in which a product's value to one user depends on its use by other users, as in the distinctive communities of people on specific social media platforms. Someone concerned about AI at one company may not have good alternatives, dissuading them from choosing a product more aligned with the public interest. Additionally, AI is often only part of consumer-facing products. Concern about AI may be outweighed by other concerns. For example, a user may appreciate the cultural and educational value of a media sharing site (such as YouTube or Instagram) even if they dislike its recommendation algorithm.

Finally, public action may tend to be primarily oriented toward how AI systems are deployed. Earlier phases of the AI system lifecycle have fewer direct ties to the public and are therefore less likely to garner public attention. For these phases of the AI lifecycle, other types of action may tend to be more important.

4.8. The Media

The media, as the term is used in this paper, refers to both professional and amateur journalists together with their diverse means of distribution, from traditional newspapers to online social media platforms. The media can play an important role in improving AI corporate governance by researching, documenting, analyzing, and drawing attention to good practices, problems, and areas for improvement. The media serves as an important link between actors internal and external to the corporation, and it plays a vital role in distilling and explaining complex technology and business detail in terms that can be understood and used by outside audiences including the public and policymakers. Indeed, media reports have been essential for the insights contained in this paper.

AI corporate governance is often in the news. Several newspapers have dedicated technology sections including *The New York Times*, *Wall Street Journal*, and *Financial Times.* Dedicated technology media sources include *The Verge*, *Wired*, and the *MIT Technology Review*. *The Markup* is specifically focused on the societal impacts of digital technology companies. All of these outlets devote extensive attention to AI corporate governance.

Media coverage has been instrumental in highlighting problems in AI corporate governance and mobilizing pressure for change. For example, the media has published several reports based on worker whistleblowing, presenting issues at AI companies that otherwise may have stayed internal to the companies (see Section 4.2). In the case of Google's participation in Project Maven, *Gizmodo* reported about the project based on leaked information [106]. The media also covered the subsequent protests by Google workers, further amplifying their concerns [107,177,178]. Google management later announced it would leave the project [108]. This example demonstrates how broad media coverage combined with employee activism can have significant influence on corporate decision-making.

Other reporting focuses on adverse societal impacts of AI. One prominent example is a ProPublica investigative report on biased outcomes of the COMPAS algorithm for assessing the risk of a person committing a future crime [179]. The report has been credited with focusing researchers' attention on fairness in machine learning [180] (p. 29). A more recent example is a *New York Times* exposé on the little-known facial recognition company Clearview [181], which was scraping personal photos uploaded online to serve as training data. The *Times* report prompted technology companies to take legal action [182] and nonprofit advocacy organizations to lobby the U.S. government to intervene [183].

Media coverage appears to be most robust for later phases of the AI system lifecycle. Later phases, such as the deployment of AI products, tend to be more publicly visible and of more direct interest to the public and other outside stakeholders. In contrast, earlier phases, such as basic research and development, are less visible and less directly relevant to stakeholders, and therefore may tend to receive less coverage. Coverage of internal

corporate activities may be impossible without whistleblowers; these activities can occur across the lifecycle but may be especially frequent during earlier phases.

Other factors can also shape trends in the media coverage of corporate AI. Coverage may tend to be greater where AI intersects with other topics of great public interest, such as racism or unemployment, or for prominent AI companies or celebrity entrepreneurs. Certain events can also draw new attention to existing practices, such as the longstanding privacy flaws of the Zoom videoconferencing platform gaining newfound attention due to the heavy use of Zoom during the COVID-19 pandemic [184]. Risky AI practices may tend to receive the greatest attention in the immediate aftermath of an incident of that type of risk, unless such incidents are too commonplace to be considered newsworthy. This can result in less coverage of emerging and extreme AI risks for which incidents have not previously occurred. (The tendency to overlook extreme risks has been dubbed "the tragedy of the uncommons" [185])

Finally, as with nonprofits, the media faces financial challenges. The business models of many media outlets have been harmed by the rise of the internet and other recent factors. This has resulted in less investigative journalism, meaning fewer resources to report on issues in AI corporate governance. Meanwhile, the AI industry is amidst a financial boom, making it difficult for the media to hire people with expertise in AI. There can even be conflicts of interest, as has been a concern since Amazon founder Jeff Bezos purchased the *Washington Post* (a concern that Bezos and the *Post* deny [186]). The media clearly has an important role in advancing AI corporate governance in the public interest, making it vital that its various challenges be overcome.

4.9. Government

Government, as the term is used here, refers to institutions with legal authority over some geographic jurisdiction, whether national (e.g., the United States), subnational (e.g., California), or supranational (e.g., the E.U.). Governments can influence AI corporate governance to promote the public interest by using various policy instruments. In the following, we consider four widely accepted categories of policy instruments: command and control regulation, market-based instruments, soft law, and information and education. We also consider procurement, which is important for the government use of AI. Our focus on these categories of instruments reflects current practices to influence the corporate governance of AI; however, more methods could be used in the future, including the role of state-owned enterprises and direct subsidies.

Command and control regulation uses binding legal rules to specify the required behavior, and enforcement measures to correct or halt non-compliant behavior [187] (p. 107). Many existing regulations are applicable to AI, although they do not address AI specifically. For example, in the E.U., AI systems must comply with existing data protection, consumer protection, and anti-discrimination laws [188] (p. 13). The GDPR contains detailed rules governing the processing of personal data using automated decision-making [189]. In this case, the person whose data are being processed can request "meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing," although whether this entails a right to explanation is disputed [34,35]. Although these rules are not explicitly about AI (automated decision-making as defined in the GDPR is not synonymous with AI), they are nonetheless applicable to many AI systems [190]. Similarly, labor law is generally not written explicitly for work in AI, but it nonetheless affects how workers and management are allowed to act, as seen for example in allegations of former Google workers that they were fired because of their labor organizing, which may have been in violation of U.S. labor law (Section 4.2) [101].

In recent years, governments around the world have started to work on AI-specific command and control regulations. Proposals have been published, among others, by the E.U. [28,188,191], China [192], and the United States [193,194]. For example, in April 2021, the European Commission published a proposal for an Artificial Intelligence Act [28], following its White Paper on AI [188] and the High-Level Expert Group on AI's Ethics

Guidelines for Trustworthy AI [191]. The new proposal follows a risk-based approach with different requirements for different levels of risk. It prohibits practices which pose unacceptable risks (e.g., social scoring by governments or systems that exploit vulnerabilities of children) and contains specific rules for high-risk systems (e.g., biometric identification systems). These rules include requirements regarding the quality of datasets used; technical documentation and record keeping; transparency and the provision of information to users; human oversight; and robustness, accuracy and cybersecurity. The proposed regulation contains very light and mostly voluntary provisions for AI systems with low or minimal risk. The vast majority of AI systems currently used in the E.U. fall into this category. The requirements for high-risk systems are command and control because they require specific behavior that will be enforced by supervisory authorities. It is worth noting that the proposal still has to be adopted by the European Parliament and the member states.

Market-based instruments affect corporate activity through economic incentives [195] (p. 22); [187] (p. 117). For example, taxes could be used to encourage safe behavior or discourage unsafe behavior, such as via tax discounts for AI systems that have been certified or audited by a third party. Civil liability could incentivize AI companies to mitigate risks from accidents or the misuse of AI systems. Subsidies could support corporate initiatives on AI in the public interest, for example, to support research and development on AI techniques that improve safety. These benefits may also support the public good insofar as they address a market failure or incentivize innovation with future public benefit. The DARPA Grand Challenge for Autonomous Vehicles is one such example, which helped catalyze private investment in the field in the early 2000s [196].

Procurement refers to government purchases of goods and services [197,198], including AI systems. For example, law enforcement agencies have recently sought to purchase facial recognition software. As discussed throughout this paper, this procurement is controversial, with many arguing that it is not in the public interest. This controversy speaks to the fact that governments face important decisions on which AI systems to procure and how to use them to best advance the public interest. Governments can additionally use procurement to influence AI corporate governance by procuring systems that meet high standards of safety and ethics. This incentivizes industry to adopt and maintain such standards. Procurement is thus, in a sense, a demand-side market-based instrument in its potential to use market incentives to advance AI corporate governance in the public interest.

As discussed in Section 4.5, soft law is the non-binding expectation of behavior that has some indirect legal basis. One specific form of soft law in which governments play a central role is co-regulation. It is worth noting that co-regulation is not necessarily a form of soft law, but the concepts are at least interconnected [40,41]. Co-regulation expands on corporate self-regulation to include some government involvement, typically to ensure enforcement [195] (p. 35). For example, the U.S. Federal Trade Commission (FTC) encourages firms to declare privacy policies, and prosecutes firms who deviate from their statements [199]. Conceivably, the FTC could also punish violations of companies' self-stated AI principles [30]. However, such enforcement has been ineffective. In 2011, Facebook agreed to a settlement with the FTC after being accused of violating its privacy policy [200], but the violations continued, most notably with the 2018 Cambridge Analytica scandal [201]. In 2019, Facebook again settled with the FTC, this time for an unprecedented USD 5 billion and stringent monitoring requirements [202]. However, even the USD 5 billion fine could be seen as simply the cost of business, given that Facebook's 2019 profit was nearly USD 18.5 billion [203], and especially if the costs can be amortized across multiple years.

Governments can also lead public information and education campaigns [187] (p. 116). A better educated public could incentivize AI companies to improve their governance, as detailed in Section 4.7. Education campaigns could also foster constructive public debate on AI ethics and safety [191] (p. 23). Education takes time, and thus is unlikely to be effective in time-critical situations, but it is otherwise found to often be a cost-effective policy option [195]. Governments can lead AI education campaigns. They can also

obtain information about AI corporations, such as by establishing information disclosure requirements as discussed above.

The efficacy of policy instruments can depend on their enforcement. This applies to command and control regulation as well as certain soft law and market-based instruments. Where enforcement is applicable, it is used to ensure compliance. Noncompliance is commonly sanctioned by fines. In extreme cases, corporate activities can be shut down and noncompliant corporate personnel can be imprisoned. A lesser punishment could see companies added to published lists of noncompliant companies, as is the practice in European financial market regulation [204]. However, in practice, governments do not always vigorously enforce compliance. Monitoring and enforcement can be expensive, and governments may not always have the resources or motivation to do so. Weak enforcement can limit the influence of government rules and regulations on AI corporate governance. Additionally, if the people responsible for complying with AI regulations disagree with or resent them—and are sufficiently empowered to act on this disagreement—then it could prompt backlash, such that the people decline to comply and may even do more of the disallowed or discouraged behavior than would occur without the regulation [59].

When selecting a policy instrument to improve the corporate governance of AI, governments need to consider a number of factors. One of these factors is the underlying regulatory approach. Most AI-specific proposals follow a risk-based approach. This approach ensures that regulation does not exceed what is necessary to achieve the underlying policy objective, as is required by the principle of proportionality in E.U. law. Apart from that, governments need to decide between regulation focused on specific technologies, such as AI, or regulation that addresses general issues that can be applicable to multiple technologies, such as privacy. Finally, governments need to decide whether the regulation should be cross-sector or for specific sectors, such as healthcare or transportation.

AI-specific policy instruments face a particular challenge in defining their scope of application [29] (pp. 359–362). There is no generally accepted definition of the term AI, and existing definitions do not meet the specific requirements for legal definitions [205]. A possible solution to this problem would be to avoid the term AI and define other properties of the system, such as certain use cases or technical approaches. This idea may be a worthy focus of future AI policy research and activity.

Policy shapes innovation [206] (p. 249), and this will be no different with AI. Regulation can impose costs on or the outright prohibition of certain types of AI research and applications, thereby limiting innovation in particular areas while making others more attractive. Meanwhile, market mechanisms and procurement may subsidize or otherwise incentivize some types of AI research and development over others. For example, law enforcement procurement of facial recognition may already be stimulating innovation in that branch of AI. Taken as a whole, then, government regulation may be expected to shape AI innovation. Poorly constructed regulation may shape, or particularly limit, innovation in ways that undermine the public interest; this is a bug to be remedied, not a feature. For example, poorly constructed regulation may place a disproportionate burden of AI-related regulatory compliance that falls on smaller companies with fewer resources.

When a government uses policy instruments, the effects are not always limited to that government's jurisdiction, a phenomenon known as regulatory diffusion. Corporations that work across jurisdictions often follow the regulations of the most stringent jurisdiction across all jurisdictions to gain the efficiencies of a single standardized compliance operation. They may even lobby other jurisdictions to adopt similar rules so as to similarly bind local competitors. Influential jurisdictions in this regard include California and the European Union, sometimes referred to as the "California Effect" [207] and the "Brussels Effect" [208]. Given that leading AI corporations are multinational, policy instruments from a wide range of jurisdictions could shape corporate governance in (their conception of) the public interest globally. Even if companies are not incentivized to comply globally, other jurisdictions may pass similar regulation, imitating the first mover. Regulations do not always diffuse,

and corporations may shift operations to jurisdictions with relatively lax regulations. Nonetheless, regulatory diffusion can increase the impacts of policy innovation.

Regulation on law enforcement use of facial recognition technology has demonstrated such regulatory imitation. Municipalities seem to have taken the lead in the United States. Following research on AI and gender and race biases [58,209], some municipalities have started to ban the use of facial recognition technology for law enforcement purposes, including San Francisco [210] and Boston [211]. Even though municipal action has set the agenda for wider action, leading to multiple bills that have been introduced in the U.S. Congress on this topic [212,213], there is not yet a regulation of facial recognition technology at the federal level. Due to the absence of federal regulation, the legal treatment of the technology is currently highly variable across the United States. In keeping with their incentives for regulatory consistency across jurisdictions, Microsoft has repeatedly called for the federal regulation of facial recognition technology [135,214,215]. Under the proposed E.U. AI regulation, all remote biometric identification of persons will be considered high-risk and subject to third party conformity assessment [28]. Certain applications for the purpose of law enforcement will be prohibited in principle with a few narrow exceptions.

Governments may not simply wait for AI policy instruments to passively diffuse; they may support institutionalized international coordination. For example, the OECD AI Principles have been adopted by 45 nations and informed similar principles agreed to by the G-20 countries [216]. The OECD AI Policy Observatory is now developing implementation guidance for the principles, aimed at both government and corporate decision-makers. International coordination is further supported by several UN initiatives [217]. Pre-existing international agreements and initiatives can shape AI corporate governance. Prominent examples include the UN Guiding Principles for Business and Human Rights and the UN Global Compact, which offer guidance for business practices to promote the public interest. Already, Google has adapted some AI systems according to this UN guidance [218]. Additionally, the OECD has published general corporate governance guidance that has been adopted into national regulation [219,220].

5. Conclusions

A wide range of actors can help improve the corporate governance of AI so as to better advance the public interest. It is not, as one might think, a matter for the exclusive attention of a narrow mix of insider corporate elites. To be sure, the opportunities may often be better for some types of actors than others. However, significant opportunities can be found for many actors both within and outside of AI corporations. Importantly, these opportunities are available even in the absence of dedicated multistakeholder institutions that are designed to invite contributions from a more diverse group. Multistakeholder institutions have an important role to play, but they are only one of many means through which diverse stakeholders can improve AI corporate governance.

Often, progress depends on coordination and collaboration across different types of actors, as illustrated in the three primary cases used throughout the paper. First, the example of Google's Project Maven shows that workers and the media can collaborate to be particularly successful in influencing management. Second, the example of law enforcement use of facial recognition technology demonstrates that novel research, activism by nonprofits, and broad media coverage can build on each other to achieve change in corporate governance. Third, the example of the publication of potentially harmful research shows management, workers, and industry consortia interacting to establish, implement, and share best practices for AI in the public interest. People in each type of actor category would do well to understand not just their own opportunities, but also the broader ecosystem of actors and their interactions. Questions of how best to pursue coordination and collaboration across actors must be resolved on a case-by-case basis, in consideration of the particulars of the issues at play and the relative roles and capabilities of different actors.

Opportunities to improve AI corporate governance are likely to change over time. For example, workers' influence may diminish over time if the expected increase in the supply of skilled AI workers outpaces demand increases for AI systems. Changes in the economic, cultural, and political significance of AI can alter the opportunities available to many types of actors, such as by shaping the political viability of government regulations. Changes in underlying AI technologies can also be impactful. For example, if there are major breakthroughs toward AI systems that can substitute for most forms of human labor or even approach an artificial general intelligence, then the corporations could end up with substantially greater economic and political clout. This could increase the importance of actions from within the companies, especially from management and investors. On the other hand, such breakthroughs could also increase public and policymaker interest in AI in ways that facilitate more extensive government activity over time. The delay in government responses to emerging technologies, often called the "pacing problem" [221], creates a clear, even if interim, role for other actors in improving AI corporate governance in the public interest as AI research and development continues.

This paper has presented a broad survey of opportunities to improve AI corporate governance across a range of actors. It is, to the best of our knowledge, the first such survey published. As a first pass through a large and complex topic, the paper's survey has been largely qualitative, and it has also not been comprehensive in its scope. We have sought to map out the overall terrain of AI corporate governance without necessarily identifying or measuring all the hills and valleys. We have focused on select larger corporations, with less attention to smaller ones. We have focused on the United States and Europe, with less attention to other parts of the world. Additionally, to at least some extent, we have covered a convenience sample of cases. The result is a broad but somewhat limited map of the AI corporate governance landscape.

One important area for future research is on evaluating the quality of opportunities to improve AI corporate governance. Specific actors may benefit from guidance on how best to focus their activities. Some actors, such as researchers and philanthropists, have opportunities to bolster the efforts of other types of actors and would benefit from guidance on which other actors are most worth supporting. (These supporting actions fall outside the scope of this paper's framework and would make for a further area for future research.) To a large extent, the quality of opportunities facing specific actors must be assessed on a case-by-case basis accounting for context and technological particulars, and therefore fall outside the scope of broad surveys such as this paper. An important activity would be to bridge the gap between the more general insights of this survey and the specific insights needed for corporate governance decision-making in the public interest for particular categories of AI policy problems and types of AI systems. Such work should include more specific conceptions of the public interest, because different conceptions can underlie different evaluative standards and generate different practical guidance.

Finally, further research could investigate how AI corporate governance may change over time, especially as companies develop increasingly capable AI systems. This paper has emphasized near-term cases in order to give its study of AI corporate governance a better empirical basis and more immediate practical value. Nonetheless, the potential for extremely large consequences from long-term AI make it a worthy subject of attention. Important questions include how near-term actions could affect long-term AI corporate governance, such as through path dependence in governance regimes, and how future actors can best position themselves to influence long-term corporate AI for the better. One good starting point may be to look more closely at the earliest phases of the AI lifecycle, especially basic research and development, on the grounds that this may be where future advanced forms of AI first appear within corporations.

As the deployment of AI systems and research and development of AI technologies continue, the role of AI corporate governance is expected to only increase over time. With it too increases the importance of experimentation and iteration to develop actors' strategies to improve the corporate governance of AI companies in the public interest. This paper has

surveyed the landscape with the aim of empowering practitioners and catalyzing necessary further research. It is only with this continued work, by the full range of actors considered here, that AI corporations may be expected to support the public interest today, tomorrow, and into the future.

Author Contributions: Conceptualization, P.C., J.S. and S.D.B.; research, P.C., J.S. and S.D.B.; analysis, P.C., J.S. and S.D.B.; writing, P.C., J.S and S.D.B.; funding acquisition, S.D.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Gordon R. Irlam Charitable Foundation.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: For helpful input on this research, we thank participants in a seminar hosted by the Global Catastrophic Risk Institute. For input on an initial draft, we are grateful to Ramiro de Avila Peres, Rosie Campbell, Alexis Carlier, Sam Clarke, Jessica Cussins-Newman, Moritz Kleinaltenkamp, Yolanda Lannquist, Joel Lehman, Jeremy Nixon, Dakota Norris, and Cullen O'Keefe. We thank Oliver Couttolenc for research assistance and feedback and McKenna Fitzgerald for assistance in manuscript formatting. Any remaining errors are the authors' alone. Views expressed here are those of the authors and do not necessarily represent the views of their employers.

Conflicts of Interest: The authors declare no conflict of interest. The paper was researched and drafted prior to Peter Cihon joining his current employer, GitHub, a subsidiary of Microsoft.

References

1. Government of France. Launch of the Global Partnership on Artificial Intelligence. 2020. Available online: https://www.gouvernement.fr/en/launch-of-the-global-partnership-on-artificial-intelligence (accessed on 11 September 2020).
2. European Commission High-Level Expert Group on AI. Policy and Investment Recommendations for Trustworthy Artificial Intelligence. 2019. European Commission Website. Available online: https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence (accessed on 11 September 2020).
3. Cath, C.; Wachter, S.; Mittelstadt, B.; Taddeo, M.; Floridi, L. Artificial intelligence and the 'good society': The US, EU, and UK approach. *Sci. Eng. Ethics* **2017**, *24*, 505–528. [CrossRef]
4. Perrault, R.; Shoham, Y.; Brynjolfsson, E.; Clark, J.; Etchemendy, J.; Grosz, B.; Lyons, T.; Manyika, J.; Mishra, S.; Niebles, J.C. *The AI Index 2019 Annual Report*; Human-Centered AI Institute, Stanford University: Stanford, CA, USA, 2019.
5. Frey, C.B.; Osborne, M.A. The future of employment: How susceptible are jobs to computerisation? *Technol. Forecast. Soc. Chang.* **2017**, *114*, 254–280. [CrossRef]
6. Baum, S.D. A Survey of Artificial General Intelligence Projects for Ethics, Risk, and Policy Working Paper 17-1. 2017. Available online: https://dx.doi.org/10.2139/ssrn.3070741 (accessed on 24 June 2021).
7. Radford, A.; Wu, J.; Amodei, D.; Amodei, D.; Clark, J.; Brundage, M.; Sutskever, I. Better language models and their implications. Available online: https://openai.com/blog/better-language-models (accessed on 11 September 2020).
8. Partnership on AI. Partnership on AI Publication Norms for Responsible AI. Available online: https://www.partnershiponai.org/case-study/publication-norms (accessed on 11 September 2020).
9. Gilson, R.J. From corporate law to corporate governance. In *The Oxford Handbook of Corporate Law and Governance*; Gordon, J.N., Ringe, W.-G., Eds.; Oxford University Press: Oxford, UK, 2016; Volume 1, pp. 3–27. ISBN 9780198743682.
10. Stout, L.A. *The Shareholder Value Myth: How Putting Shareholders First Harms Investors, Corporations, and the Public*; Berrett-Koehler Publishers: San Francisco, CA, USA, 2012; ISBN 9781605098135.
11. Business Roundtable. Business Roundtable Redefines the Purpose of a Corporation to Promote 'An Economy That Serves All Americans'. 2019. Available online: https://www.businessroundtable.org/business-roundtable-redefines-the-purpose-of-a-corporation-to-promote-an-economy-that-serves-all-americans (accessed on 11 September 2020).
12. Freeman, R.E. *Strategic Management: A Stakeholder Approach*; Pitman: Boston, MA, USA, 1984; ISBN 9780273019138.
13. Raymond, M.; DeNardis, L. Multistakeholderism: Anatomy of an inchoate global institution. *Int. Theory* **2015**, *7*, 572–616. [CrossRef]
14. Freeman, E.; Martin, K.; Parmar, B. Stakeholder capitalism. *J. Bus. Ethics* **2007**, *74*, 303–314. [CrossRef]
15. Legg, S.; Hutter, M. Universal intelligence: A definition of machine intelligence. *Minds Mach.* **2007**, *17*, 391–444. [CrossRef]
16. Marcus, G.; Davis, E. *Rebooting AI: Building Artificial Intelligence We Can Trust*; Pantheon Books: New York, NY, USA, 2019; ISBN 9780525566045.

17. McCorduck, P. *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*; 25th Anniversary Update; A.K. Peters Ltd.: Natick, MA, USA, 2004; ISBN 9781568812052.
18. OECD. *Scoping the OECD AI Principles: Deliberations of the Expert Group on Artificial Intelligence at the OECD (AIGO)*; OECD Digital Economy Papers No. 291; OECD: Paris, France, 2015.
19. Monks, R.A.G.; Minow, N. *Corporate Governance*, 5th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2011; ISBN 9780470972595.
20. Gordon, J.N.; Ringe, W.-G. *The Oxford Handbook of Corporate Law and Governance; Oxford Handbooks*, 1st ed.; Oxford University Press: Oxford, UK, 2018; ISBN 9780198743682.
21. Crawford, K.; Dobbe, R.; Dryer, T.; Fried, G.; Green, B.; Kaziunas, E.; Kak, A.; Mathur, V.; McElroy, E.; Sánchez, A.N.; et al. *AI Now 2019 Report*; AI Now Institute: New York, NY, USA, 2019.
22. Metcalf, J.; Moss, E.; Boyd, D. Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics. *Soc. Res. Int. Q.* **2019**, *82*, 449–476.
23. World Economic Forum. Empowering AI Leadership. 2020. Available online: https://spark.adobe.com/page/RsXNkZANwMLEf (accessed on 11 September 2020).
24. Dafoe, A. *AI Governance: A Research Agenda*; Centre for the Governance of AI, Future of Humanity Institute, University of Oxford: Oxford, UK, 2017.
25. Calo, R. Artificial intelligence policy: A primer and roadmap. *UC Davis Law Rev.* **2017**, *51*, 399–435.
26. Calo, R. *The Case for a Federal Robotics Commission*; Brookings Institute: Washington, DC, USA, 2014; Available online: https://www.brookings.edu/research/the-case-for-a-federal-robotics-commission (accessed on 11 September 2020).
27. Floridi, L.; Cowls, J.; Beltrametti, M.; Chatila, R.; Chazerand, P.; Dignum, V.; Luetge, C.; Madelin, R.; Pagallo, U.; Rossi, F.; et al. AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds Mach.* **2018**, *28*, 689–707. [CrossRef]
28. European Commission. *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COM(2021) 206 Final)*; European Commission: Brussels, Belgium, 2021.
29. Scherer, M.U. Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies. *Harv. J. Law Technol.* **2016**, *29*, 354–400. [CrossRef]
30. Wallach, W.; Marchant, G.E. An agile ethical/legal model for the international and national governance of AI and robotics. In Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, New Orleans, LA, USA, 2–3 February 2018; ACM: New York, NY, USA, 2018.
31. Erdelyi, O.J.; Goldsmith, J. Regulating artificial intelligence proposal for a global solution. In Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, New Orleans, LA, USA, 2–3 February 2018; ACM: New York, NY, USA, 2018; pp. 95–101.
32. Cihon, P.; Maas, M.M.; Kemp, L. Should artificial intelligence governance be centralised? Design lessons from history. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, New York, NY, USA, 7–9 February 2020; ACM: New York, NY USA, 2020; pp. 228–234.
33. Clark, J.; Hadfield, G.K. Regulatory markets for AI safety. In Proceedings of the 2019 Safe Machine Learning Workshop at ICLR, New Orleans, LA, USA, 6 May 2019; ICLR: La Jolla, CA, USA, 2019.
34. Wachter, S.; Mittelstadt, B.; Floridi, L. Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *Int. Data Priv. Law* **2017**, *7*, 76–99. [CrossRef]
35. Goodman, B.; Flaxman, S. European Union regulations on algorithmic decision-making and a "right to explanation". *AI Mag.* **2017**, *38*, 50–57. [CrossRef]
36. Smuha, N.A. From a "Race to AI" to a "Race to AI Regulation"—Regulatory Competition for Artificial Intelligence. 2019. Available online: https://dx.doi.org/10.2139/ssrn.3501410 (accessed on 11 September 2020).
37. Thelisson, E.; Padh, K.; Celis, E.L. Regulatory Mechanisms and Algorithms towards Trust in AI/ML. 2017. Available online: https://www.researchgate.net/publication/318913104_Regulatory_Mechanisms_and_Algorithms_towards_TruTr_in_AIML (accessed on 11 September 2020).
38. Stix, C. *A Survey of the European Union's Artificial Intelligence Ecosystem*; Lverhulme Centre for the Future of Intelligence, University of Cambridge: Cambridge, UK, 2019.
39. Wagner, B.; Rozgonyi, K.; Sekwenz, M.-T.; Cobbe, J.; Singh, J. Regulating Transparency? Facebook, Twitter and the German Network Enforcement Act. In Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '20), Barcelona, Spain, 27–30 January 2020; ACM: New York, NY, USA, 2020; pp. 261–271.
40. Senden, L. Soft law, self-regulation and co-regulation in European law: Where do they meet? *EJCL* **2005**, *9*, 1–27.
41. Marsden, C.T. *Internet Co-Regulation European Law, Regulatory Governance and Legitimacy in Cyberspace*; Cambridge University Press: Cambridge, UK, 2011. [CrossRef]
42. Kaminski, M.E. Binary governance: Lessons from the GDPR's approach to algorithmic accountability. *South. Calif. Law Rev.* **2019**, *92*, 1529–1616. [CrossRef]
43. Pagallo, U. The middle-out approach: Assessing models of legal governance in data protection, artificial intelligence, and the web of data. *Theory Pract. Legis.* **2019**, *7*, 1–25. [CrossRef]
44. Zeitlin, J. *Extending Experimentalist Governance? The European Union and Transnational Regulation*; Oxford University Press: Oxford, UK, 2015; ISBN 9780198724506.

45. Marchant, G.; Lindor, R. The coming collision between autonomous vehicles and the liability System. *St. Clara Law Rev.* **2012**, *52*, 1321–1340.
46. LeValley, D. Autonomous vehicle liability—Application of common carrier liability. *Seattle Univ. Law Rev. Supra* **2013**, *36*, 5–26.
47. Zohn, J.R. When robots attack: How should the law handle self-driving cars that cause damages. *J. Law Technol. Policy* **2015**, *2015*, 461–485.
48. Bathaee, Y. The artificial intelligence black box and the failure of intent and causation. *Harv. J. Law Technol.* **2018**, *31*, 889–938.
49. Lohmann, M.F. Ein europäisches Roboterrecht—Überfällig oder überflüssig? *ZRP* **2017**, *6*, 168–171.
50. Cauffman, C. Robo-liability: The European Union in search of the best way to deal with liability for damage caused by artificial intelligence. *Maastricht J. Eur. Comp. Law* **2018**, *25*, 527–532. [CrossRef]
51. European Parliament. European Parliament Resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)). 2017. Available online: https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.pdf (accessed on 11 September 2020).
52. Expert Group on Liability and New Technologies—New Technologies Formation. *Liability for Artificial Intelligence and Other Emerging Digital Technologies*; European Commission: Brussels, Belgium, 2019; ISBN 9789276129592.
53. European Commission. *Report from the Commission to the European Parliament, the Council, and the European Economic and Social Committee (COM(2020) 324 final)*; European Commission: Brussels, Belgium, 2020.
54. Denga, M. Deliktische Haftung für künstliche Intelligenz—Warum die Verschuldenshaftung des BGB auch künftig die bessere Schadensausgleichsordnung bedeutet. *CR* **2018**, 69–78. [CrossRef]
55. Borges, G. Rechtliche Rahmenbedingungen für autonome Systeme. *NJW* **2018**, *40*, 977–982.
56. Graf von Westphalen, F. Haftungsfragen beim Einsatz Künstlicher Intelligenz in Ergänzung der Produkthaftungs-RL 85/374/EWG. *ZIP* **2020**, *40*, 889–895.
57. White, T.N.; Baum, S.D. Liability for present and future robotics technology. In *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*; Lin, P., Abney, K., Jenkins, R., Eds.; Oxford University Press: Oxford, UK, 2017; Volume 1, pp. 66–79.
58. Buolamwini, J.; Gebru, T. Gender shades: Intersectional accuracy disparities in commercial gender classification. In Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '18), New York, NY, USA, 23–24 February 2018; ACM: New York, NY, USA, 2018; pp. 77–91.
59. Baum, S.D. On the promotion of safe and socially beneficial artificial intelligence. *AI Soc.* **2017**, *32*, 543–551. [CrossRef]
60. Belfield, H. Activism by the AI community: Analysing recent achievements and future prospects. In Proceedings of the 2020 AAAI/ACM Conference on AI, Ethics and Society, New York, NY, USA, 7–8 February 2020; ACM: New York, NY, USA, 2020; pp. 15–21.
61. Askell, A.; Brundage, M.; Hadfield, G. The Role of Cooperation in Responsible AI Development. 2019. Available online: http://arxiv.org/abs/1907.04534 (accessed on 11 September 2020).
62. Solaiman, I.; Brundage, M.; Clark, J.; Askell, A.; Herbert-Voss, A.; Wu, J.; Radford, A.; Krueger, G.; Kim, J.W.; Kreps, S.; et al. Release Strategies and the Social Impacts of Language Models. OpenAI. 2019. Available online: http://arxiv.org/abs/1908.09203 (accessed on 11 September 2020).
63. Cihon, P. *Standards for AI Governance: International Standards to Enable Global Coordination in AI Research & Development*; Future of Humanity Institute, University of Oxford: Oxford, UK, 2019.
64. Baum, S.D. Superintelligence skepticism as a political tool. *Information* **2018**, *9*, 209. [CrossRef]
65. Baum, S.D. Countering Superintelligence Misinformation. *Information* **2018**, *9*, 244. [CrossRef]
66. O'Keefe, C.; Cihon, P.; Garfinkel, B.; Flynn, C.; Leung, J.; Dafoe, A. The Windfall Clause: Distributing the benefits of AI for the common good. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, New York, NY, USA, 7–8 February 2020; ACM: New York, NY, USA, 2020; pp. 327–331.
67. Avin, S.; Gruetzemacher, R.; Fox, J. Exploring AI futures through role play. In Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, New Orleans, LA, USA, 2–3 February 2018; ACM: New York, NY, USA, 2018; pp. 8–14.
68. Ballard, S.; Calo, R. Taking futures seriously: Forecasting as method in robotics law and policy. In Proceedings of the 2019 We Robot Conference, We Robot, Miami, FL, USA, 12–13 April 2019.
69. Hume, K.; LaPlante, A. Managing Bias and Risk at Every Step of the AI-Building Process. *Harvard Business Review*, 30 October 2019. Available online: https://hbr.org/2019/10/managing-bias-and-risk-at-every-step-of-the-ai-building-process (accessed on 11 September 2020).
70. Tiell, S. Create an Ethics Committee to Keep Your AI Initiative in Check. *Harvard Business Review*, 15 November 2019. Available online: https://hbr.org/2019/11/create-an-ethics-committee-to-keep-your-ai-initiative-in-check (accessed on 11 September 2020).
71. Chamorro-Premuzic, T.; Polli, F.; Dattner, B. Building Ethical AI for Talent Management. *Harvard Business Review*, 21 November 2019. Available online: https://hbr.org/2019/11/building-ethical-ai-for-talent-management (accessed on 11 September 2020).
72. Fountaine, T.; McCarthy, B.; Saleh, T. Building the AI-Powered Organization. *Harvard Business Review*, 1 July 2019. Available online: https://hbr.org/2019/07/building-the-ai-powered-organization (accessed on 11 September 2020).
73. Abbasi, A.; Kitchens, B.; Ahmad, F. The Risks of AutoML and How to Avoid Them. *Harvard Business Review*, 24 October 2019. Available online: https://hbr.org/2019/10/the-risks-of-automl-and-how-to-avoid-them (accessed on 11 September 2020).

74. Hao, K. Establishing an AI Code of Ethics Will be Harder than People Think. *MIT Technology Review*, 21 October 2018. Available online: https://www.technologyreview.com/2018/10/21/139647/establishing-an-ai-code-of-ethics-will-be-harder-than-people-think (accessed on 11 September 2020).
75. Hao, K. In 2020, Let's Stop AI Ethics-Washing and Actually do Something. *MIT Technology Review*, 27 December 2019. Available online: https://www.technologyreview.com/2019/12/27/57/ai-ethics-washing-time-to-act (accessed on 11 September 2020).
76. Burkhardt, R.; Hohn, N.; Wigley, C. Leading Your Organization to Responsible AI. *McKinsey Co.*, 2 May 2019. Available online: https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/leading-your-organization-to-responsible-ai (accessed on 11 September 2020).
77. Cheatham, B.; Javanmardian, K.; Samandari, H. Confronting the Risks of Artificial Intelligence. *McKinsey Co.*, 26 April 2019. Available online: https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/confronting-the-risks-of-artificial-intelligence (accessed on 11 September 2020).
78. Ransbotham, S.; Khodabandeh, S.; Fehling, R.; LaFountain, B.; Kiron, D. *Winning with AI: Pioneers Combine Strategy, Organizational Behavior, and Technology*; MIT Sloan Management Review and Boston Consulting Group: Boston, MA, USA, 2019.
79. PWC. A Practical Guide to Responsible Artificial Intelligence (AI). 2019. Available online: https://www.pwc.com/gx/en/issues/data-and-analytics/artificial-intelligence/what-is-responsible-ai/responsible-ai-practical-guide.pdf (accessed on 11 September 2020).
80. Ernst & Young Global Limited. *How Do You Teach AI the Value of Trust?* Report No. 03880-183Gbl; Ernst & Young Global Limited: London, UK, 2018; Available online: https://www.ey.com/en_us/digital/how-do-you-teach-ai-the-value-of-trust (accessed on 11 September 2020).
81. KPMG. Controlling AI: The Imperative for Transparency and Explainability. 2019. Available online: https://advisory.kpmg.us/content/dam/advisory/en/pdfs/kpmg-controlling-ai.pdf (accessed on 11 September 2020).
82. Deloitte. AI and Risk Management. Available online: https://www2.deloitte.com/gr/en/pages/financial-services/articles/gx-ai-and-risk-management.html (accessed on 11 September 2020).
83. Accenture. Building Data and Ethics Committees. 2019. Available online: https://www.accenture.com/_acnmedia/PDF-107/Accenture-AI-And-Data-Ethics-Committee-Report-11.pdf (accessed on 11 September 2020).
84. Pye, L.W.; Verba, S. *Political Culture and Political Development*; Princeton University Press: Princeton, NJ, USA, 1965; p. 7.
85. Jobin, A.; Ienca, M.; Vayena, E. The global landscape of AI ethics guidelines. *Nat. Mach. Intell.* **2019**, *1*, 389–399. [CrossRef]
86. Morley, J.; Floridi, L.; Kinsey, L.; Elhalal, A. From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Sci. Eng. Ethics* **2020**, *26*, 2141–2168. [CrossRef]
87. Gibney, E. The battle for ethical AI at the world's biggest machine-learning conference. *Nature* **2020**, *577*, 609. [CrossRef]
88. Raji, I.D.; Smart, A.; White, R.N.; Mitchell, M.; Gebru, T.; Hutchinson, B.; Smith-Loud, J.; Theron, D.; Barnes, P. Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19), Atlanta, GA, USA, 29–31 January 2019; ACM: New York, NY, USA, 2019; pp. 220–229.
89. Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J.W.; Wallach, H.; Daumé, H., III; Crawford, K. Datasheets for Datasets. 2020. Available online: http://arxiv.org/abs/1803.09010 (accessed on 11 September 2020).
90. Mitchell, M.; Wu, S.; Zaldivar, A.; Barnes, P.; Vasserman, L.; Hutchinson, B.; Spitzer, E.; Raji, I.D.; Gebru, T. Model Cards for Model Reporting. In Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19), Atlanta, GA, USA, 29–31 January; ACM: New York, NY, USA, 2019; pp. 220–229.
91. OpenAI Charter. OpenAI. Available online: https://openai.com/charter (accessed on 11 September 2020).
92. Brockman, G.; Sutskever, I.; OpenAI LP. *OpenAI*. 11 March 2019. Available online: https://openai.com/blog/openai-lp (accessed on 11 September 2020).
93. Smith, R. The Future of Face Matching at Axon and AI Ethics Board Report. *Axon*, 27 June 2019. Available online: https://www.axon.com/news/ai-ethics-board-report (accessed on 11 September 2020).
94. Piper, K. Exclusive: Google cancels AI ethics board in response to outcry. *Vox*, 4 April 2019. Available online: https://www.vox.com/future-perfect/2019/4/4/18295933/google-cancels-ai-ethics-board (accessed on 11 September 2020).
95. Google. *Google's Approach to IT Security: A Google White Paper*; Google: Mountain View, CA, USA, 2012; Available online: https://static.googleusercontent.com/media/1.9.22.221/en//enterprise/pdf/whygoogle/google-common-security-whitepaper.pdf (accessed on 11 September 2020).
96. Cooper, D. Towards a model of safety culture. *Saf. Sci.* **2000**, *36*, 111–136. [CrossRef]
97. Kinstler, L. Ethicists were hired to save tech's soul. Will anyone let them? *Protocol*, 5 February 2020. Available online: https://www.protocol.com/ethics-silicon-valley (accessed on 11 September 2020).
98. Hao, K. The messy, secretive reality behind OpenAI's bid to save the world. *MIT Technology Review*, 17 February 2020. Available online: https://www.technologyreview.com/2020/02/17/844721/ai-openai-moonshot-elon-musk-sam-altman-greg-brockman-messy-secretive-reality (accessed on 11 September 2020).
99. Johnson, K. NeurIPS requires AI researchers to account for societal impact and financial conflicts of interest. *VentureBeat*, 24 February 2020. Available online: https://venturebeat.com/2020/02/24/neurips-requires-ai-researchers-to-account-for-societal-impact-and-financial-conflicts-of-interest (accessed on 11 September 2020).

100. Simonite, T. What really happened when Google ousted Timnit Gebru. *Wired*. 8 June 2021. Available online: https://www.wired.com/story/google-timnit-gebru-ai-what-really-happened (accessed on 15 June 2021).
101. De Vynck, G.; Bergen, M.; Gallagher, R.; Barr, A. Google fires four employees, citing data-security violations. *Bloomberg Law*. 25 November 2019. Available online: https://www.bloomberg.com/news/articles/2019-11-25/google-fires-four-employees-citing-data-security-violations (accessed on 11 September 2020).
102. Nicas, J. Google tries to corral its staff after ugly internal debates. *The New York Times*. 23 August 2019. Available online: https://www.nytimes.com/2019/08/23/technology/google-culture-rules.html (accessed on 11 September 2020).
103. Conger, K.; Wakabayashi, D. Google fires 4 workers active in labor organizing. *The New York Times*. 25 November 2019. Available online: https://www.nytimes.com/2019/11/25/technology/google-fires-workers.html (accessed on 11 September 2020).
104. Shoham, Y.; Perrault, R.; Brynjolfsson, E.; Clark, J.; Manyika, J.; Niebles, J.C.; Lyons, T.; Etchemendy, J.; Grosz, B.; Bauer, Z. *The AI Index 2018 Annual Report*; Human-Centered AI Institute, Stanford University: Stanford, CA, USA, 2018.
105. Dutton, T. *Building an AI World: Report on National and Regional AI Strategies*; CIFAR: Ontario, Canada, 2018; Available online: https://www.cifar.ca/cifarnews/2018/12/06/building-an-ai-world-report-on-national-and-regional-ai-strategies (accessed on 11 September 2020).
106. Cameron, D.; Conger, K. Google Is Helping the Pentagon Build AI for Drones. *Gizmodo*. 6 March 2018. Available online: https://gizmodo.com/google-is-helping-the-pentagon-build-ai-for-drones-1823464533 (accessed on 11 September 2020).
107. Shane, S.; Wakabayashi, D. 'The business of war': Google employees protest work for the Pentagon. *The New York Times*. 4 April 2018. Available online: https://www.nytimes.com/2018/04/04/technology/google-letter-ceo-pentagon-project.html (accessed on 11 September 2020).
108. Wakabayashi, D.; Shane, S. Google will not renew Pentagon contract that upset employees. *The New York Times*. 1 June 2018. Available online: https://www.nytimes.com/2018/06/01/technology/google-pentagon-project-maven.html (accessed on 11 September 2020).
109. Gallagher, R. Google plans to launch censored search engine in China, leaked documents reveal. *The Intercept*. 1 August 2018. Available online: https://theintercept.com/2018/08/01/google-china-search-engine-censorship (accessed on 11 September 2020).
110. Google Employees Against Dragonfly. We are Google employees. Google must drop Dragonfly. *Medium*. 27 November 2018. Available online: https://medium.com/@googlersagainstdragonfly/we-are-google-employees-google-must-drop-dragonfly-4c8a30c5e5eb (accessed on 10 September 2020).
111. Alba, D. A Google VP told the US Senate the company has "terminated" the Chinese search app Dragonfly. *BuzzFeed News*. 6 July 2019. Available online: https://www.buzzfeednews.com/article/daveyalba/google-project-dragonfly-terminated-senate-hearing (accessed on 11 September 2020).
112. Wakabayashi, D.; Benner, K. How Google protected Andy Rubin, the 'Father of Android'. *The New York Times*. 25 October 2018. Available online: https://www.nytimes.com/2018/10/25/technology/google-sexual-harassment-andy-rubin.html (accessed on 11 September 2020).
113. Stapleton, C.; Gupta, T.; Whittaker, M.; O'Neil-Hart, C.; Parker, S.; Anderson, E.; Gaber, A. We're the organizers of the Google walkout. Here are our demands. *The Cut*. 1 November 2018. Available online: https://www.thecut.com/2018/11/google-walkout-organizers-explain-demands.html (accessed on 11 September 2020).
114. Google Walkout for Real Change. #GoogleWalkout update: Collective action works, but we need to keep working. *Medium*. 11 November 2018. Available online: https://medium.com/@GoogleWalkout/googlewalkout-update-collective-action-works-but-we-need-to-keep-wworkin-b17f673ad513 (accessed on 11 September 2020).
115. Employees of Microsoft. An open letter to Microsoft: Don't bid on the US military's Project JEDI. *Medium*. 16 October 2018. Available online: https://medium.com/s/story/an-open-letter-to-microsoft-dont-bid-on-the-us-military-s-project-jedi-7279338b7132 (accessed on 10 September 2020).
116. Smith, B. Technology and the US military. *Microsoft*. 26 October 2018. Available online: https://blogs.microsoft.com/on-the-issues/2018/10/26/technology-and-the-us-military (accessed on 11 September 2020).
117. An Amazon Employee. I'm an Amazon employee. My company shouldn't sell facial recognition tech to police. *Medium*. 16 October 2018. Available online: https://medium.com/@amazon_employee/im-an-amazon-employee-my-company-shouldn-t-sell-facial-recognition-tech-to-police-36b5fde934ac (accessed on 11 September 2020).
118. Merchant, B. 6000 Amazon employees, including a VP and directors, are now calling on Jeff Bezos to stop automating oil extraction. *Gizmodo*. 1 April 2019. Available online: https://gizmodo.com/6-000-amazon-employees-including-a-vp-and-directors-n-1834001079 (accessed on 11 September 2020).
119. Grewal, J.; Serafeim, G.; Yoon, A. *Shareholder Activism on Sustainability Issues*; Harvard Business School Working Paper, No. 17-003; Harvard Business School: Boston, MA, USA, 2016; Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2805512 (accessed on 11 September 2020).
120. Ben-Amar, W.; Chang, M.; McIlkenny, P. Board gender diversity and corporate response to sustainability initiatives: Evidence from the Carbon Disclosure Project. *J. Bus. Ethics* **2017**, *142*, 369–383. [CrossRef]
121. Sharton, B.R.; Stegmaier, G.M.; Procter, G. Breaches in the boardroom: What directors and officers can do to reduce the risk of personal liability for data security breaches. *Reuters*. Available online: https://legal.thomsonreuters.com/en/insights/articles/board-liability-reduce-risk-for-data-security-breaches (accessed on 11 September 2020).

122. Sawyer, M. *Annual Review and Analysis of 2019 U.S. Shareholder Activism*; Sullivan & Cromwell LLP: New York, NY, USA, 2019; Available online: https://www.sullcrom.com/siteFiles/Publications/2019ShareholderActivismAnnualReport.pdf (accessed on 11 September 2020).
123. U.S. Securities Exchange Commission. How to Read a 10-K. 2011. Available online: https://www.sec.gov/fast-answers/answersreada10khtm.html (accessed on 11 September 2020).
124. Chow, C.; Frame, K.; Likhtman, S.; Spooner, N.; Wong, J. *Investors' Expectations on Responsible Artificial Intelligence and Data Governance*; Hermes Investment Management: London, UK, 2019; Available online: https://www.hermes-investment.com/eos-insight/eos/investors-expectations-on-responsible-artificial-intelligence-and-data-governance (accessed on 11 September 2020).
125. Hermes EOS calls on Alphabet to lead responsible A.I. practice. In *U.S. Securities Exchange Commission Website*; 17 June 2019. Available online: https://www.sec.gov/Archives/edgar/data/1013143/000108514619001758/hermes-alphabet061919.htm (accessed on 11 September 2020).
126. Lahoti, S. Google rejects all 13 shareholder proposals at its annual meeting, despite protesting workers. *Packt Hub*. 20 June 2019. Available online: https://hub.packtpub.com/google-rejects-all-13-shareholder-proposals-at-its-annual-meeting-despite-protesting-workers (accessed on 11 September 2020).
127. Aten, J. Google has a date with shareholders today and they are telling the company it's time for a break up. *Inc*. 19 June 2019. Available online: https://www.inc.com/jason-aten/google-has-a-date-with-shareholders-today-they-are-telling-company-its-time-for-a-break-up.html (accessed on 11 September 2020).
128. Amazon. Proxy Statement: 2019 Annual Meeting of Shareholders. 2019. Available online: https://s2.q4cdn.com/299287126/files/doc_financials/proxy/2019-Proxy-Statement.pdf (accessed on 11 September 2020).
129. Amazon. Notice of 2020 Annual Meeting of Shareholders & Proxy Statement. 2020. Available online: https://s2.q4cdn.com/299287126/files/doc_financials/2020/ar/updated/2020-Proxy-Statement.pdf (accessed on 11 September 2020).
130. Dastin, J.; Kerber, R.U.S. blocks Amazon efforts to stop shareholder votes on facial recognition. *Reuters*. 5 April 2019. Available online: https://www.reuters.com/article/us-amazon-com-facial-recognition-idUSKCN1RG32N (accessed on 11 September 2020).
131. Strubell, E.; Ganesh, A.; McCallum, A. Energy and policy considerations for deep learning in NLP. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; ACL: Stroudsburg, PA, USA, 2019; pp. 3645–3650.
132. DiMaggio, P.J.; Powell, W.W. The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *Am. Sociol. Rev.* **1983**, *48*, 147–160. [CrossRef]
133. IBM. IBM CEO's Letter to Congress on Racial Justice Reforms. 2020. Available online: https://www.ibm.com/blogs/policy/facial-recognition-sunset-racial-justice-reforms (accessed on 11 September 2020).
134. Amazon. We Are Implementing a One-Year Moratorium on Police Use of Rekognition. 2020. Available online: https://blog.aboutamazon.com/policy/we-are-implementing-a-one-year-moratorium-on-police-use-of-rekognition (accessed on 11 September 2020).
135. Washington Post Live (@postlive) Washington Post Live on Twitter: "Microsoft president @BradSmi says the company does not sell facial recognition software to police depts. in the U.S. today and will not sell the tools to police until there is a national law in place 'grounded in human rights.' #postlive https://t.co/lwxBLjrtZL". *Twitter*. 11 June 2020. Available online: https://twitter.com/postlive/status/1271116509625020417 (accessed on 11 September 2020).
136. Google. Celebrity Recognition. Cloud Vision API. Available online: https://cloud.google.com/vision/docs/celebrity-recognition (accessed on 11 September 2020).
137. Menn, J. Microsoft turned down facial-recognition sales on human rights concerns. *Reuters*. 4 April 2019. Available online: https://www.reuters.com/article/us-microsoft-ai-idUSKCN1RS2FV (accessed on 11 September 2020).
138. Article One Advisors. Case Studies: Microsoft. Available online: https://www.articleoneadvisors.com/microsoft (accessed on 11 September 2020).
139. Nicas, J. Atlanta asks Google whether it targeted Black homeless people. *The New York Times*. 4 October 2019. Available online: https://www.nytimes.com/2019/10/04/technology/google-facial-recognition-atlanta-homeless.html (accessed on 11 September 2020).
140. Kumar, R.S.S.; Nagle, F. The Case for AI Insurance. *Harvard Business Review*. 29 April 2020. Available online: https://hbr.org/2020/04/the-case-for-ai-insurance (accessed on 11 September 2020).
141. Franke, U. The cyber insurance market in Sweden. *Comput. Secur.* **2017**, *68*, 130–144. [CrossRef]
142. Kosik, A. FedEx asks the Washington Redskins to change their name after pressure from investor groups. *CNN*. 3 July 2020. Available online: https://www.cnn.com/2020/07/02/business/fedex-washington-redskins/index.html (accessed on 11 September 2020).
143. Partnership on AI. Meet the Partners. Available online: https://www.partnershiponai.org/partners (accessed on 11 September 2020).
144. IEEE SA. IEEE Standards Association Membership. Available online: https://standards.ieee.org/content/ieee-standards/en/about/membership (accessed on 11 September 2020).
145. Leibowicz, C.; Adler, S.; Eckersley, P. When is it appropriate to publish high-stakes AI research? *Partnership on AI*. 2 April 2019. Available online: https://www.partnershiponai.org/when-is-it-appropriate-to-publish-high-stakes-ai-research (accessed on 11 September 2020).

146. Socher, R. Introducing a conditional transformer language model for controllable generation. *Salesforce*. 11 September 2019. Available online: https://blog.einstein.ai/introducing-a-conditional-transformer-language-model-for-controllable-generation (accessed on 11 September 2020).
147. Keskar, N.S.; McCann, B.; Varshney, L.R.; Xiong, C.; Socher, R. CTRL: A Conditional Transformer Language Model for Controllable Generation. 2019. Available online: http://arxiv.org/abs/1909.05858 (accessed on 11 September 2020).
148. Anandwala, R.; Cassagnol, D. CTA launches first-ever industry-led standard for AI in health care. *Consumer Technology Association*. 25 February 2020. Available online: https://cta.tech/Resources/Newsroom/Media-Releases/2020/February/CTA-Launches-First-Ever-Industry-Led-Standard (accessed on 11 September 2020).
149. Black, J.; Hopper, M.; Band, C. Making a success of principles-based regulation. *Law Financ. Mark. Rev.* **2007**, *1*, 191–206. [CrossRef]
150. Cihon, P.; Gutierrez, G.M.; Kee, S.; Kleinaltenkamp, M.J.; Voigt, T. *Why Certify? Increasing Adoption of the Proposed EU Cybersecurity Certification Framework*; Judge Business School, University of Cambridge: Cambridge, UK, 2018.
151. Meyer, T. Soft law as delegation. *Fordham Int. Law J.* **2009**, *32*, 888–942.
152. Marchant, G.E. "Soft law" governance of artificial intelligence. *AI Pulse*. 25 January 2019. Available online: https://aipulse.org/soft-law-governance-of-artificial-intelligence (accessed on 11 September 2020).
153. Google. *Perspectives on Issues in AI Governance*; Google: Mountain View, CA, USA, 2019; Available online: https://ai.google/static/documents/perspectives-on-issues-in-ai-governance.pdf (accessed on 11 September 2020).
154. Oreskes, N.; Conway, E.M. *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*; Bloomsbury Press: New York, NY, USA, 2010; ISBN 9781596916104.
155. Ali, M.; Sapiezynski, P.; Bogen, M.; Korolova, A.; Mislove, A.; Rieke, A. Discrimination through optimization: How Facebook's ad delivery can lead to skewed outcomes. In Proceedings of the ACM on Human-Computer Interaction, Lake Buena Vista, FL, USA, 26–28 July 2019; ACM: New York, NY, USA, 2019; Volume 3, pp. 1–30.
156. Ranking Digital Rights. *2019 RDR Corporate Accountability Index*; Ranking Digital Rights: Budapest, Hungary, 2019; Available online: https://rankingdigitalrights.org/index2019/assets/static/download/RDRindex2019report.pdf (accessed on 11 September 2020).
157. Gebhart, G. *Who Has Your Back? Censorship Edition 2019*; Electronic Frontier Foundation: San Francisco, CA, USA, 2019; Available online: https://www.eff.org/wp/who-has-your-back-2019 (accessed on 11 September 2020).
158. AI Now Institute. Publications. Available online: https://ainowinstitute.org/reports.html (accessed on 11 September 2020).
159. ACLU. Petition: Amazon: Get Out of the Surveillance Business. Available online: https://action.aclu.org/petition/amazon-stop-selling-surveillance (accessed on 11 September 2020).
160. Snow, J. Amazon's Face recognition falsely matched 28 members of Congress with mugshots. *ACLU*. 26 July 2018. Available online: https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28 (accessed on 11 September 2020).
161. ACLU National. An open letter to Amazon shareholders. *Medium*. 20 May 2019. Available online: https://medium.com/aclu/an-open-letter-to-amazon-shareholders-374f4fb84e98 (accessed on 11 September 2020).
162. Mullins, B.; Nicas, J. Paying professors: Inside Google's academic influence campaign. *Wall Street Journal*. 14 July 2017. Available online: https://www.wsj.com/articles/paying-professors-inside-googles-academic-influence-campaign-1499785286 (accessed on 11 September 2020).
163. Taplin, J. Google's disturbing influence over think tanks. *The New York Times*. 30 August 2017. Available online: https://www.nytimes.com/2017/08/30/opinion/google-influence-think-tanks.html (accessed on 11 September 2020).
164. Tully, S.M.; Winer, R.S. The role of the beneficiary in willingness to pay for socially responsible products: A meta-analysis. *Soc. Responsib. Prod. Supply Chain Manag. EJournal* **2014**. [CrossRef]
165. Bijker, W.E. , *Hughes, T.P., Pinch, T. The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology, Anniversary ed.*; MIT Press: Cambridge, MA, USA, 2012; ISBN 9780262517607.
166. Business Insider Intelligence. The messaging apps report: Messaging apps are now bigger than social networks. *Business Insider*. 16 September 2016. Available online: https://www.businessinsider.com/the-messaging-app-report-2015-11 (accessed on 11 September 2020).
167. Legge, J.S., Jr.; Durant, R.F. Public opinion, risk assessment, and biotechnology: Lessons from attitudes toward genetically modified foods in the European Union. *Rev. Policy Res.* **2010**, *27*, 59–76. [CrossRef]
168. Wiener, J.B.; Rogers, M.D. Comparing precaution in the United States and Europe. *J. Risk Res.* **2002**, *5*, 317–349. [CrossRef]
169. Parker, K.; Horowitz, J.M.; Anderson, M. Majorities across racial, ethnic groups express support for the Black Lives Matter movement. *Pew Research Center*. 12 June 2020. Available online: https://www.pewsocialtrends.org/2020/06/12/amid-protests-majorities-across-racial-and-ethnic-groups-express-support-for-the-black-lives-matter-movement (accessed on 11 September 2020).
170. Brewster, T. The many ways Google Glass users risk breaking British privacy laws. *Forbes*. 30 June 2014. Available online: https://www.forbes.com/sites/thomasbrewster/2014/06/30/the-many-ways-google-glass-users-risk-breaking-bribrit-privacy-laws (accessed on 11 September 2020).
171. Google. Google Glass. Available online: https://www.google.com/glass/start (accessed on 11 September 2020).

172. Simonite, T. When it comes to gorillas, Google Photos remains blind. *Wired*. 11 January 2018. Available online: https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind (accessed on 11 September 2020).
173. Vogel, D. *The Market for Virtue: The Potential and Limits of Corporate Social Responsibility*; Brookings Institution Press: Washington, DC, USA, 2006; ISBN 9780815790761.
174. Porter, M.E.; Kramer, M.R. Strategy and society: The link between competitive advantage and corporate social responsibility. *Harvard Business Review*. 1 December 2006. Available online: https://hbr.org/2006/12/strategy-and-society-the-link-between-competitive-advantage-and-corporate-social-resresponsibil (accessed on 11 September 2020).
175. Elements of AI. Elements of AI—Join the movement! Available online: http://www.elementsofai.com/eu2019fi (accessed on 11 September 2020).
176. Baum, S.D. Medium-term artificial intelligence and society. *Information* **2020**, *11*, 290. [CrossRef]
177. Deahl, D. Google employees demand the company pull out of Pentagon AI project. *The Verge*. 4 April 2018. Available online: https://www.theverge.com/2018/4/4/17199818/google-pentagon-project-maven-pull-out-letter-ceo-sundar-pichpi (accessed on 11 September 2020).
178. Griffith, E. Google won't renew controversial Pentagon AI project. *Wired*. 1 June 2018. Available online: https://www.wired.com/story/google-wont-renew-controversial-pentagon-ai-project (accessed on 11 September 2020).
179. Angwin, J.; Larson, J.; Mattu, S.; Kirchner, L. Machine bias. *ProPublica*. 23 May 2016. Available online: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (accessed on 11 September 2020).
180. Partnership on AI. *Report on Algorithmic Risk Assessment Tools in the U.S. Criminal Justice System*; Partnership on AI: San Francisco, CA, USA; Available online: https://www.partnershiponai.org/report-on-machine-learning-in-risk-assessment-tools-in-the-u-s-criminal-justice-system (accessed on 11 September 2020).
181. Hill, K. The secretive company that might end privacy as we know it. *The New York Times*. 1 January 2020. Available online: https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html (accessed on 11 September 2020).
182. Statt, N. Controversial facial recognition firm Clearview AI facing legal claims after damning NYT report. *The Verge*. 24 January 2020. Available online: https://www.theverge.com/2020/1/24/21079354/clearview-ai-nypd-terrorism-suspect-false-claims-facial-recognireco (accessed on 11 September 2020).
183. Alianza Naciónal de Campesinas; Algorithmic Justice League; American-Arab Anti-Discrimination Committee; American Friends Service Committee; Black and Brown Activism Defense Collective; Campaign for a Commercial-Free Childhood; Center for Digital Democracy; Coalition for Humane Immigrant Rights; Color of Change; Constitutional Alliance; et al. *PCLOB Letter of Suspension of Facial Recognition Technology*; Electronic Privacy Information Center: Washington, DC, USA, 2020; Available online: https://epic.org/privacy/facerecognition/PCLOB-Letter-FRT-Suspension.pdf (accessed on 11 September 2020).
184. Paul, K. Zoom releases security updates in response to "Zoom-bombings". *The Guardian*. 23 April 2020. Available online: http://www.theguardian.com/technology/2020/apr/23/zoom-update-security-encryption-bombing (accessed on 11 September 2020).
185. Wiener, J.B. The tragedy of the uncommons: On the politics of apocalypse. *Glob. Policy* **2016**, *7*, 67–80. [CrossRef]
186. Wieczner, J. How Jeff Bezos reacts to 'negative' Amazon articles in the Washington Post. *Fortune*. 27 October 2017. Available online: https://fortune.com/2017/10/27/amazon-jeff-bezos-washington-post (accessed on 11 September 2020).
187. European Commission. *Better Regulation "Toolbox"*; European Commission: Brussels, Belguim, 2017.
188. European Commission. *White Paper on Artificial Intelligence—A European Approach to Excellence and Trust*; European Commission: Brussels, Belgium, 2020.
189. Wachter, S.; Mittelstadt, B.; Russell, C. Counterfactual explanations without opening the Black Box: Automated decisions and the GDPR. *Harv. J. Law Technol.* **2018**, *31*. [CrossRef]
190. Sartor, G.; European Parliament; European Parliamentary Research Service; Scientific Foresight Unit. *The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence: Study*; European Parliamentary Research Service: Brussels, Belgium, 2020; ISBN 9789284667710.
191. Independent High-Level Expert Group on Artificial Intelligence. *Ethics Guidelines for Trustworthy AI*; Report B-1049; European Commission: Brussels, Belgium, 2019.
192. Webster, G.; Creemers, R.; Triolo, P.; Kania, E. Full translation: China's "New Generation Artificial Intelligence Development Plan". *New American*. 1 August 2017. Available online: http://newamerica.org/cybersecurity-initiative/digichina/blog/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017 (accessed on 11 September 2020).
193. The White House. Executive Order on Maintaining American Leadership in Artificial Intelligence. In *The White House*; 11 February 2019. Available online: https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence (accessed on 11 September 2020).
194. Vought, R.T. Memorandum for the Heads of Executive Departments and Agencies. In *The White House*; 2020. Available online: https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf (accessed on 11 September 2020).
195. Hepburn, G. *Alternatives to Traditional Regulation*; OECD: Paris, France, 2006.
196. DARPA. The Grand Challenge for Autonomous Vehicles. Available online: https://www.darpa.mil/about-us/timeline/-grand-challenge-for-autonomous-vehicles (accessed on 11 September 2020).
197. Edler, J.; Georghiou, L. Public procurement and innovation—Resurrecting the demand side. *Res. Policy* **2007**, *36*, 949–963. [CrossRef]

198. Edquist, C.; Zabala-Iturriagagoitia, J.M. Public procurement for innovation as mission-oriented innovation policy. *Res. Policy* **2012**, *41*, 1757–1769. [CrossRef]
199. Hetcher, S. The FTC as internet privacy norm entrepreneur. *Vanderbilt Law Rev.* **2000**, *53*, 2041–2062. [CrossRef]
200. U.S. Federal Trade Commission. Facebook Settles FTC Charges That It Deceived Consumers by Failing to Keep Privacy Promises. 2011. Available online: https://www.ftc.gov/news-events/press-releases/2011/11/facebook-settles-ftc-charges-it-deceived-consumers-failing-keep (accessed on 11 September 2020).
201. Confessore, N. Cambridge Analytica and Facebook: The scandal and the fallout so far. *The New York Times*. 4 April 2018. Available online: https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html (accessed on 11 September 2020).
202. Fair, L. FTC's $5 billion Facebook settlement: Record-breaking and history-making. In *U.S. Federal Trade Commission*; 424 July 2019. Available online: https://www.ftc.gov/news-events/blogs/business-blog/2019/07/ftcs-5-billion-facebook-settlement-record-breaking-history (accessed on 11 September 2020).
203. Facebook Investor Relations. *Facebook Reports Fourth Quarter and Full Year 2019 Results*; Facebook: Menlo Park, CA, USA, 2020; Available online: https://investor.fb.com/investor-news/press-release-details/2020/Facebook-Reports-Fourth-Quarter-and-Full-Year-2019-Results/default.aspx (accessed on 11 September 2020).
204. European Parliament; Council of the European Union. Regulation (EU) No 596/2014 of the European Parliament and of the Council of 16 April 2014 on market abuse (market abuse regulation) and repealing Directive 2003/6/EC of the European Parliament and of the Council and Commission Directives 2003/124/EC, 2003/125/EC and 2004/72/EC Text with EEA relevance. *OJL* **2014**, *173*, 1–61.
205. Schuett, J. A Legal Definition of AI. *arXiv* **2019**, arXiv:1909.01095. [CrossRef]
206. Blind, K.; Petersen, S.S.; Riillo, C.A.F. The impact of standards and regulation on innovation in uncertain markets. *Res. Policy* **2017**, *46*, 249–264. [CrossRef]
207. Vogel, D. *Trading Up: Consumer and Environmental Regulation in a Global Economy*; Harvard University Press: Cambridge, MA, USA, 1995; ISBN 9780674900837.
208. Bradford, A. *The Brussels Effect: How the European Union Rules the World*; Oxford University Press: New York, NY, USA, 2020; ISBN 9780190088583.
209. West, S.M.; Whittaker, M.; Crawford, K. *Discriminating Systems: Gender, Race, and Power in AI*; AI Now Institute: New York, NY, USA, 2019; p. 33.
210. Conger, K.; Fausset, R.; Kovaleski, S.F. San Francisco bans facial recognition technology. *The New York Times*. 14 May 2019. Available online: https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html (accessed on 11 September 2020).
211. Johnson, K. Boston bans facial recognition due to concern about racial bias. *VentureBeat*. 24 June 2020. Available online: https://venturebeat.com/2020/06/24/boston-bans-facial-recognition-due-to-concern-about-racial-bias (accessed on 11 September 2020).
212. Blunt, R. S.847—116th Congress (2019–2020): Commercial Facial Recognition Privacy Act of 2019. 14 March 2019.
213. Merkley, J. S.3284—116th Congress (2019–2020): Ethical Use of Facial Recognition Act. 12 February 2020.
214. Smith, B. Facial recognition technology: The need for public regulation and corporate responsibility. *Microsoft*. 13 July 2018. Available online: https://blogs.microsoft.com/on-the-issues/2018/07/13/facial-recognition-technology-the-need-for-public-regulation-and-corporate-responsibility (accessed on 11 September 2020).
215. Smith, B. Facial recognition: It's time for action. *Microsoft*. 6 December 2018. Available online: https://blogs.microsoft.com/on-the-issues/2018/12/06/facial-recognition-its-time-for-action (accessed on 11 September 2020).
216. OECD. Principles on Artificial Intelligence. Available online: https://www.oecd.org/going-digital/ai/principles (accessed on 11 September 2020).
217. Butcher, J.; Beridze, I. What is the state of artificial intelligence governance globally? *RUSI J.* **2019**, *164*, 88–96. [CrossRef]
218. BSR. Google Celebrity Recognition API Human Rights Assessment Executive Summary. Available online: https://www.bsr.org/reports/BSR-Google-CR-API-HRIA-Executive-Summary.pdf (accessed on 11 September 2020).
219. OECD. *Due Diligence Guidance for Responsible Business Conduct*; OECD Publishing: Paris, France, 2018.
220. OECD. *Due Diligence Guidance for Responsible Supply Chains of Minerals from Conflict-Affected and High-Risk Areas*, 3rd ed.; OECD Publishing: Paris, France, 2016; ISBN 9789264252387.
221. Marchant, G.E.; Allenby, B.R.; Herkert, J.R. *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight the Pacing Problem*; The International Library of Ethics, Law and Technology; Springer: Amsterdam, The Netherlands, 2011; Volume 7, ISBN 9789400713567.