Check for updates

Business & Society
1–45
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00076503221080959
journals.sagepub.com/home/bas
SAGE

An Eye for Artificial Intelligence: Insights Into the Governance of Artificial Intelligence and Vision for Future Research

Deepika Chhillar[1] and Ruth V. Aguilera[2]

Abstract

In this 60th anniversary of *Business & Society* essay, we seek to make three main contributions at the intersection of governance and artificial intelligence (AI). First, we aim to illuminate some of the deeper social, legal, organizational, and democratic challenges of rising AI adoption and resulting algorithmic power by reviewing AI research through a governance lens. Second, we propose an AI governance framework that aims to better assess AI challenges as well as how different governance modalities (architecture, laws, norms, and market) can support AI. At the heart of our framework lies the governance forces that apply to institutions, organizations, and individuals, who ultimately provide, regulate, and use AI decision-making. We discuss how businesses may harness AI's economic power through governance solutions without creating or amplifying societal biases and inequalities. Third, as part of our section on future research, we identify a set of governance trade-offs in AI adoption, suggest future research avenues to conceptually strengthen research on the governance of AI, and lay out key policy recommendations.

Keywords

artificial intelligence, big data, governance, institutions, machine learning

[1]University of Illinois at Urbana–Champaign, USA
[2]Northeastern University, Boston, MA, USA

Corresponding Author:
Deepika Chhillar, Gies College of Business, University of Illinois at Urbana–Champaign, 1206 S. Sixth Street, Champaign, IL 61820, USA.
Email: deepika4@illinois.edu

The exponential growth in data and computational power has fueled a rapid advancement of data-driven and computation-based technologies, such as artificial intelligence (AI). While AI has delivered many benefits, particularly in machine learning (ML) applications (such as image and speech recognition devices), it has also introduced new ethical, legal, and governance challenges. These include risks of unintended discrimination, privacy concerns, imbalanced property rights division, as well as algorithmic biases, to name a few. To celebrate the 60th anniversary of *Business & Society*, we seek to discuss the importance of governing AI, a conversation pertinent to businesses and society.

The use of AI adds economic value to firms, markets, and nations while radically transforming their operations by utilizing algorithms to analyze large quantities of data, generate insights from those data, enhance customer service, and predict individual and organizational outcomes (Brynjolfsson & McAfee, 2017; Felten et al., 2021; Flyverbom et al., 2019; Kellogg et al., 2020). AI refers to a set of big-data-based technologies that seek to simulate human-like thinking and decision-making in machines. These traits include but are not limited to knowledge, reasoning, problem-solving, perception, learning, and planning. AI is viewed as the "apparatus of data-technology-algorithms" (Alaimo & Kallinikos, 2020). As such, the majority of the literature discusses the two applications of AI: pattern recognition and classification algorithms. For our purposes, we focus on the three components that constitute AI (as a technological system)—big data (henceforth, data), algorithms, and machine learning. There exist multiple definitions of AI, with new ones being added each year as the field continues to evolve. We synthesize some of these definitions in Table 1, and define other related concepts and terms used in this review in Table 2.

AI-adoption by public and private societal actors trigger significant structural and institutional changes, which include shaping the legal environment (e.g., entrepreneurs filling in the vacuum within laws and regulations), setting industry practices (e.g., standard settings by market leaders in technology), giving rise to emerging organizational forms (e.g., platforms), creating new types of labor (e.g., algorithmic auditors), and emerging leadership responsibilities (e.g., shifting authority regimes and new professional roles). In addition, organizations play a crucial role in both, creating and responding to novel institutional processes that result from the adoption and proliferation of AI.

Research has identified several beneficial societal and organizational implications of employing algorithmic decision-making. For example,

Table 1. Existing Definitions of AI.

Article	Definition of AI
Alaimo & Kallinikos (2020) OS	The apparatus of data technologies—algorithms
Bolander (2019) JoMG	[AI] is almost always about building machines—computers or robots—that can perform tasks that otherwise only humans have been able lay chess, drive a car, do medical diagnosis, or engage in a dialogue.
Brock & Von Wangenheim (2019) CMR	Artificial Intelligence . . . is intended to make computers do things, that when done by people, are described as having indicated intelligence
Buhmann et al. (2020) JBE	Self-learning algorithms are a set of rules defined not by programmers but by algorithmically produced rules of learning: "The internal decision logic of the algorithm is altered as it 'learns' on training data" (Burrell, 2016, p. 5).
Chalmers et al. (2020) ENTP	Artificial intelligence is defined as "a system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation" (Kaplan & Haenlein, 2019, p. 17)
Choudhury et al. (2020) SMJ	Artificial intelligence is defined as "the capability of a machine to imitate intelligent behavior"
Felten et al. (2021) SMJ	AI refers to computer software that relies on highly sophisticated algorithmic techniques to find patterns in data and make predictions about the future. Because narrow AI algorithms "learn" from existing data to improve performance, these techniques are often referred to as "machine learning."
Glikson & Woolley (2020) AMA	A new generation of technologies capable of interacting with the environment by (a) gathering information from outside (including from natural language) or from other computer systems; (b) interpreting this information, recognizing patterns, inducing rules, or predicting events; (c) generating results, answering questions, or giving instructions to other systems; and (d) evaluating the results of their actions and improving their decision systems to achieve specific objectives
Haenlein & Kaplan (2019) CMR	A system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation.

(continued)

Table 1. (continued)

Article	Definition of AI
Huang et al. (2019) CMR	The ability to learn from various types of data and learn from a massive amount of data (i.e., big data) and update thoughts or actions is what makes us consider a machine to be intelligent
Kellogg et al. (2020) AMA	Define and work with algorithmic technologies. Defined in emerging social science usage as computer-programmed procedures that transform input data into desired outputs in ways that tend to be more encompassing, instantaneous, interactive, and opaque than previous
Kumar et al. (2019) CMR	AI refers to the broad idea that computers, through the use of software and algorithms, think and perform tasks like humans
Lévesque et al. (2020) ENTP	Merriam-Webster's dictionary defines artificial intelligence as "a branch of computer science dealing with the simulation of intelligent behavior in computers" and "the capability of a machine to imitate intelligent human behavior"
Overgoor et al. (2019) CMR	One classic textbook in the field by Russell and Norvig defines artificial intelligence as the study of the "general principles of rational agents and on components for constructing them."
Shrestha et al. (2019) CMR	The ability to learn without being explicitly programmed
Tambe et al. (2019) CMR	AI conventionally refers to a broad class of technologies that allow a computer to perform tasks that normally require human cognition, including adaptive decision-making

Note. AI = artificial intelligence; OS = Organization Studies; JoMG = Journal of Management and Governance; CMR = *California Management Review*; JBE = Journal of Business Ethics; ENTP = Entrepreneurship Theory and Practice; SMJ = Strategic Management Journal; AMA = Academy of Management Annals.

Brynjolfsson et al. (2019) show that AI has improved translation quality (for product listing titles) and economic efficiency on platforms for international trade. Similarly, Blohm and colleagues (2020) compare the investment returns of ML algorithms with those of actual investors in early-stage venture finance to find that AI can help improve decision-making in early-stage investing. In the domain of marketing, researchers celebrate the advantages

Table 2. Definitions of Keywords.

Term	Definition
Algorithm	A finite set of computational steps
Algorithmic decision-making	Algorithmic decision-making refers the use of descriptive, prescriptive, or predictive algorithms in the decision-making process
AI	The capability of using algorithms to enable human-like thinking and behavior in machines
Augmented teams	Teams where machines and humans are integrated to complement each other
Big data	The output of large-scale data mining
Big Tech	A name given to the five largest and most dominant companies in the information technology industry of the United States—namely, Google, Apple, Facebook, Amazon, and Microsoft
Data capitalism	The capacity of a select few individuals or firms to collect, store, and process large amounts of user-generated data
Data mining	The technological ability to collect, store, and process large amounts of raw data
Deep learning	A subset of machine learning algorithms that learns without human supervision, drawing from data that are both labeled and unlabeled
Machine learning	A subfield of AI which enables learning from observations (data) and experience (repeated training)
Neural networks	Algorithms that aim to mimic the way the human brain (biological neurons) operates
Strong AI	Generation of AI (futuristic) where AI will be able to achieve or even surpass human intelligence in all tasks
Surveillance capitalism	Firms or state operating outside of privacy laws and collecting free data on their citizens and users, respectively
Weak AI	Generation of AI when it can perform only one specific task but lacks the general characteristics of the human brain

Note. AI = artificial intelligence.

of personalized and targeted marketing with AI (Kumar et al., 2019; Overgoor et al., 2019). There is ample proof of AI's success in the past decade, and it is clear that many tasks are better done by AI agents than by humans—for example, automatic credit card fraud detection, credit scoring and loan approvals, detecting illness in medical images, forecasting weather patterns, and statistical analyses. However, despite claims that AI-based algorithms may be the most promising opportunity for human and organizational growth (Heukamp, 2020; Phan et al., 2017), algorithms' propensity to behave in inaccurate or unintended ways (biasedness), and their potential inscrutability to their creator (opacity), coupled with weak accountability represents a threat and danger to business and society, at large.

Governance refers to the rules and procedures that hold organizations accountable to their members and to their external stakeholders and broader society. The governance of AI-powered technologies is essential and yet complex because of AI's significant impact, as noted above. In this article, we review existing management research on AI from a governance perspective and offer a framework to examine how governance can support sustainable AI-adoption by businesses and society. We then discuss the AI challenges studied across various research themes, highlight AI governance's role in mitigating such challenges, and propose a governance framework that management scholars and policy-makers can use to make sense of this emerging yet rapidly growing research topic.

We contribute to existing research in business and society by highlighting the governance challenges of organizational AI adoption and their effects on individuals and society. In our review, we identify societal challenges emanating from ineffective governance across the multiple dimensions of AI, ranging from micro-processes related to trust in machine thinking to macro-level governmental surveillance of citizens. One of the key conclusions that emerged from our literature review is the conceptualization of governance as a trade-off concept with respect to other societal goals—in that more of one dimension directly limits another dimension because they are paradoxical. We discuss five governance trade-offs: governance versus innovation, reforming versus strengthening a surveillant state, distributed versus concentrated power; algorithm efficiency versus fair data practices, and biased AI versus societal bias, and present these as governance dilemmas in AI-adoption by businesses and society.

To address these AI governance gaps with a comprehensive approach, we draw on Lessig's (1998) New Chicago School theory to showcase one such wholesome governance approach involving socio-economic forces of laws, the market, social norms, and the architecture of the technological ecosystem. We conclude by offering suggestions for future empirical research and theory-building on the governance of AI. Our article most significantly contributes to the literature by elucidating how organizations and their governance may harness AI's power without creating or amplifying societal inequalities.

Our article is organized in three main sections. We begin by providing a background of why business and society would be served well by paying attention to the governance of AI, what it means to govern AI, and discuss how the literature has attempted to understand governance of AI thus far. Second, we describe our research design, which led us to identify three broad research themes into which the AI literature can be categorized and examine the role of governance in each of these themes. Once we have assessed the governance role in existing research, we propose a governance framework that can push forward our understanding of AI and its challenges. We close by suggesting fruitful areas of future research in the field of governance of AI, with special attention given to how existing organizational theories can help us advance knowledge in the governance of AI.

Is There a Need for More Governance?

AI in Business and Society

Strategic industries, such as health care, transportation, and energy, are becoming increasingly intelligent, efficient, and accurate due to AI applications such as robotics and autonomous vehicles, computer vision, virtual agents, and ML. Society has benefited from AI solutions such as algorithmic scans in the health care sector, tracking forestland for environmental conservation through computer vision, recognizing fraudulent card transactions in the finance sector, auto-conversion of speech to text, automatic spam detection, and many more such applications. Intelligent machines are also revolutionizing the field of caregiving, for instance, by assisting seniors (Kaplan & Haenlein, 2019; Platt, 2020) and physically challenged citizens (Potier, 2019). Moreover, better forecasting algorithms can increase consumer welfare (Miklós-Thal & Tucker, 2019; Purdy, 2020). In short, there is barely a shadow of a doubt that AI has, and will continue to have, a tremendous influence on business and society.

However, like a real engine running on oil, an AI engine running on data can produce digital fuselage in the virtual environment. Moreover, these toxins can be absorbed not just by AI users or beneficiaries, but by all members of society—children, adolescents, and grown-ups (Zwitter, 2014). To illustrate, we will focus our discussion on two broad societal challenges posed by AI that are intensified by weak governance safeguards. The first challenge is an increasing yet unbridled socio-economic power of AI-based platforms over other stakeholders in the AI ecosystem. A second challenge that AI adoption brings is surveillance capitalism which can create and perpetuate structural inequalities in society.

Intractable Socio-Economic Power of Digital Platforms

Digital platforms are "digital systems that facilitate communications, interactions, and innovations to support economic transactions and social activities" (Chen et al., 2021, p. 1306). They include some of the most dominant and successful firms in recent years, such as Amazon, Apple, Facebook, Twitter, YouTube, Uber, Airbnb, Google, Spotify, and so on. If left to self-govern, AI firms and digital platforms are likely to create negative spillovers with rising privacy breaches, health problems in a generation addicted to screens (Girish, 2020), and even an infringement of fundamental rights (Aizenberg & van den Hoven, 2020; Elkin-Koren, 2020; Sunstein, 2018) with cases such as the Cambridge Analytica scandal (Kang & Frenkel, 2018). Tech-giants have unparalleled access to user data. In this regard, West (2019) warns that users have to make a trade-off between their desires to form meaningful social relations on digital platforms and interest in protecting their personal data. Any attempts to harness AI firms' control over users' data have proven ineffective thus far, with cases of increasing recidivism (Singer & Isaac, 2020; Stempel, 2020). Part of the governance challenge is that the fines levied by the courts account for a minuscule percentage of the firms' revenues and as such, these punishments often do not deter firms from behaving unethically. A related problem is the lack of formal and informal governing rules around the definitions of unethical virtual behavior.

Research (specifically in *Business & Society*) has highlighted this concern regarding the changing power dynamics in the virtual world, particularly given the gatekeeper position of search engines. For instance, Chenou and Radu (2019) as well as Whelan (2019) analyze the role of private intermediaries in the global governance of internet-based technologies. They concur that there are a few dominating firms' growing social and political monopolistic power, as well as on the need to govern the underlying digital processes, data resources, and technical infrastructure to hold such firms accountable. Another key socio-economic concern with big digital platforms is their anti-competitive behavior in acquiring smaller firms, thereby warping the AI ecosystem in the focal firm's favor (Romm et al., 2020; Vergne, 2020; West, 2019). The few very large tech platform firms not only add existing users to their own user base, but also acquire the human capital to further innovate, thereby monopolizing the labor and product markets. Such acquisitions are likely to leave employees with fewer alternatives for employment and thus subject them to vulnerabilities in the labor market. Thus, a key question facing regulators is to figure out how to govern private players that started as innocuous businesses but are now becoming regulatory entrepreneurs (Colback, 2020; Rietveld et al., 2020) in their market domains.

Surveillance Capitalism

Surveillance capitalism involves firms or states operating outside of privacy laws and collecting free data on their users and citizens, respectively. Zuboff (2019) defines surveillance capitalism as "a new economic order that claims human experience as free raw material for . . . hidden commercial practices of extraction, prediction, and . . . behavioral modification" (p. 8). The main risks from unrestricted and ungoverned surveillance are that users may be monitored without consent and may lose autonomy over their personal decisions, owing to subtle behavioral tracking and nudges for modification. Another major threat from AI that enables surveillance capitalism is the de facto creation of user profiles (without consent). Such an auto-generated profile is called a shadow profile. Recent articles (Andrew & Baker, 2019; Etzioni, 2018; West, 2019) explore the tension between national security and individual rights, highlighting the need for developing regulatory structures, processes, and mechanisms to limit the growing social power of nations and select firms. Data on its citizens in the unregulated hands of powerful actors can be employed for socially undesirable purposes. A recent example is the Chinese government's intentional discrimination against Uighur Muslims using facial recognition technology integrated into Chinese surveillance cameras (Mozur, 2019). Therefore, surveillance risk is a challenge for society when governments and private firms monitor individuals, and access personal or non-personal information about them, with the intent to store and process these data systematically for economic or political gain. These risks related to surveillance show that there is a need to govern how these data are collected, handled, and stored to ensure that human rights are protected.

Research Design

To identify the articles to include in our literature review, we conducted a multi-step comprehensive search consisting of three phases: identification, screening, and assessment. The overall process is illustrated in Figure 1. First, with 2,000 as our starting year, we searched for the terms "artificial intelligence," "algorithm," "big data," and "machine learning" in author-supplied abstracts in the 20 top journals in management, ethics, and strategy journals as identified by FT50,[1] adding the remaining Academy of Management journals, *Business & Society* and *California Management Review*.[2] Our search yielded more than 350 articles. In the second phase, we read the abstracts and removed articles that were not peer-reviewed[3] (e.g., editorials, book reviews, and introductions to special issues). This step left us

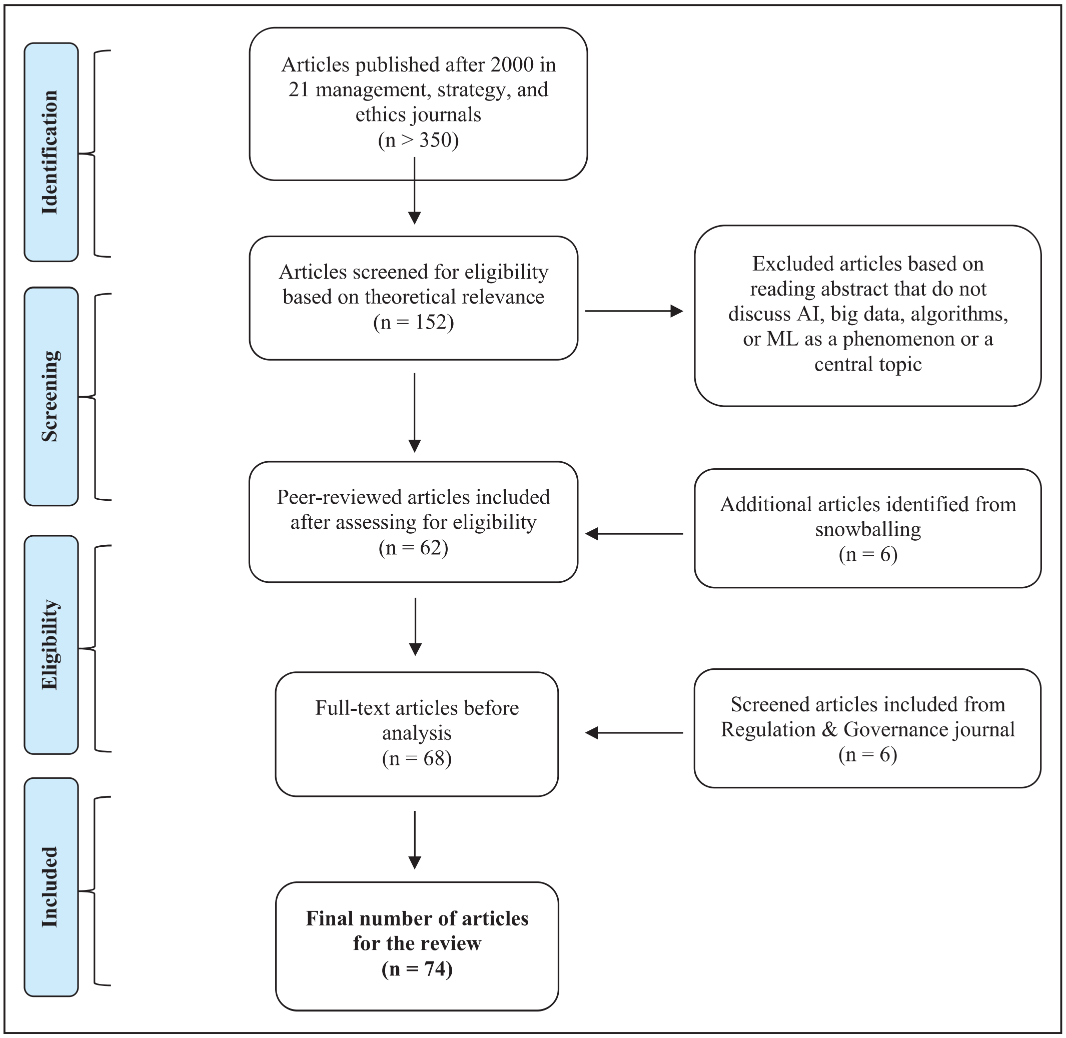

Figure 1. Literature review process.
Note. AI = artificial intelligence; ML = machine learning.

with 152 articles that we read. We eliminated from our sample those articles that did not focus on big data, ML, algorithms, or AI as a phenomenon. We also removed from our sample articles that employed algorithmic and computational methods or big data techniques in their methodology but did not discuss any individual, organizational, or societal implications of AI. This process resulted in 62 relevant articles. In the third and final phase, we did a forward and backward citation search for any additional articles that might be relevant for our review, which led us to extend our search to the *Regulation & Governance* journal because its focus aligns with our research theme on governance. Our search yielded 12 new articles. Our final sample contains 74 peer-reviewed articles, as summarized in Table A in the Online Appendix.

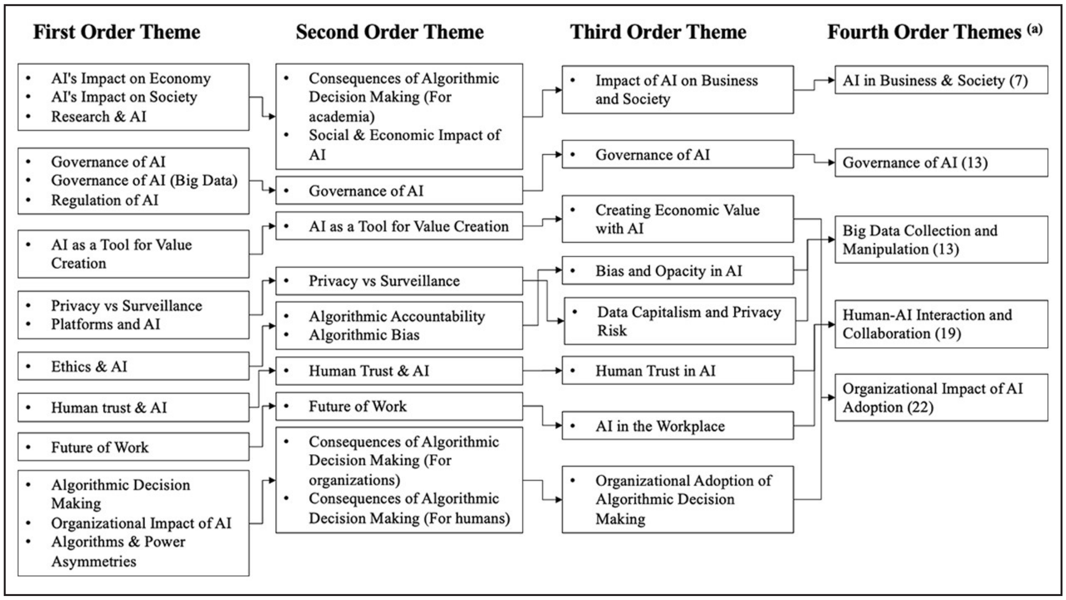

Figure 2. Recursive categorization process for review themes.
Note. AI = artificial intelligence.
[a]Indicates number of articles in each of fourth-order themes.

We thoroughly read and analyzed these 74 articles to group them into conceptual themes. Our analysis was iterative as we re-grouped the literature to categorize from 15 first-order themes into 11 second-order, to 8 third-order, and finally, 5 fourth-order themes. We constructed first-order themes based on the core arguments of each research article. When there were two or more central arguments in the article, we placed the article based on the article's main research implications and findings. Figure 2 visually represents the categorization process we followed.

We categorize the core ideas in each of the four research themes in a sequence that goes from inputs, processes, and outputs of adopting AI-based technology (as illustrated in Figure 3). In the input stage, we discuss the components that go into the ML/AI model (i.e., the data fed to the AI or ML model). In the process section, we examine the algorithms used for learning by the AI model, and in the output, we review the repercussions of using AI applications. We view these stages as inter-dependent.

Among the five themes identified in the fourth order, AI in Business and Society presents the overall message that AI is shaping our social and informational view while deeply altering businesses and society, as we discussed in the preceding section. The remaining four broad themes are Governance of AI, Big Data Collection and Manipulation, Human–AI Collaboration, and Organizational Impact of AI Adoption.

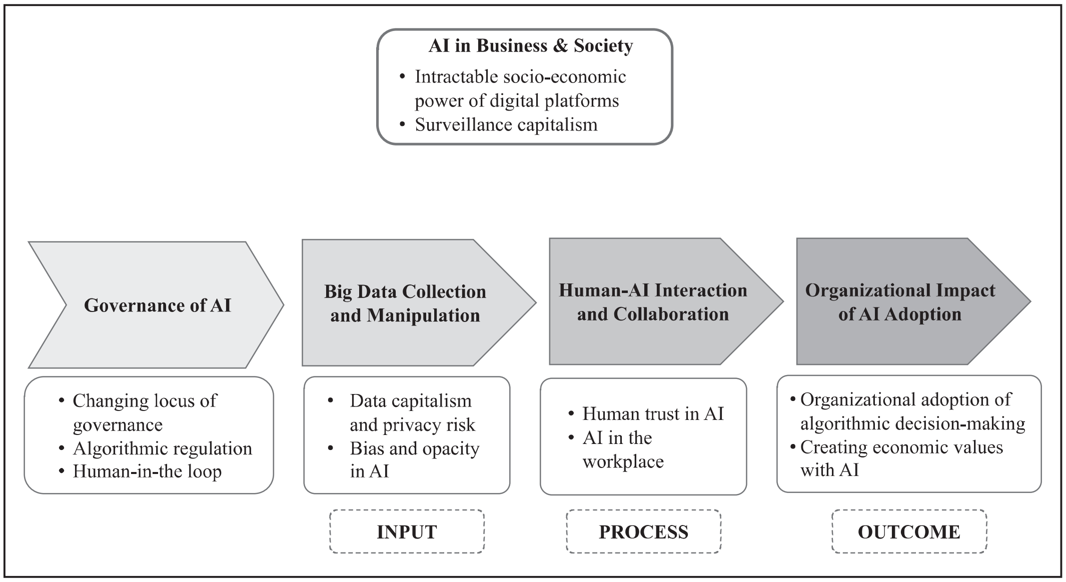

Figure 3. Outline of research themes.
Note. AI = artificial intelligence.

Literature Review

Theme: Governance of AI

Disruptive technologies within the digital field such as IoT (internet of things), big data analytics, and AI have been rapidly marching into regulatory vacuums, and existing literature has established that the norms, processes, and regulations governing AI technologies are weak and ambiguous (Arthur & Owen, 2019; Chenou & Radu, 2019; Gellert, 2020; Vergne, 2020). Because of weak regulatory governance and ambiguous industry standards, innovators and early adopters of technologies often become regulatory entrepreneurs, exposing our society to multiple risks, including but not limited to unintended externalities or negative spillovers, opportunistic actions of organizational agents, and a lack of personal and public accountability (Chen et al., 2021; Whelan, 2019; Yeung, 2018). We build upon governance insights in political science and public administration research to define the governance of AI as the structures, processes, mechanisms, and strategies that lead to the production and implementation of formal and informal rules to direct and regulate the use of AI and enforce its accountability (Levi-Faur, 2012). In this section, we review the existing literature that directly examines the theme of governance of AI and highlight some of the limitations and open areas for future research.

In our review of the AI literature, we find that scholars have stressed the need to regulate and govern platforms, organizations, industries, and sectors disrupted by shifts in business models, employment roles, and modes of value creation. We identify three main conceptual approaches that scholars have focused on: changing locus of governance, algorithmic regulation, and the human-in-the-loop approach to governance. The first set of studies explores the *changing locus of governance*. Scholars make it largely clear that governance forces from public and private players will interact to form a complete governance solution. In particular, Chenou and Radu (2019) argue in favor of hybridization of governance, whereby "public authorities de facto delegate certain types of adjudication to private actors" (p. 75), specifically in the digital sector. Several other scholars share this understanding. For instance, A. Newman (2012) notes that "governance of data privacy includes a complex web of regulation, self-regulation, and technology at the national and transnational levels, which interact to manage how personal information may be used and shared in and across modern societies" (p. 599) and adds that formal regulatory rules are the backbone of current data privacy efforts. Existing research, however, does not lay out the points of intersection or a framework of interacting regulatory forces for formulating an effective solution for the governance of AI.

A second salient trend is *algorithmic regulation* or regulation of algorithm-based decision-making. Yeung (2018) classifies algorithmic regulation as either reactive ("instruments and techniques employed by public administrators to oversee and control a set of actors and their activities") or preemptive ("policy initiatives that have sought to influence the exercise of enforcement discretion strategically in the service of institutional goals") (pp. 510–511). The author shows that there are core legal and constitutional principles at stake besides informational privacy, such as the principles of transparency and accountability, due process, and the rule of law. Eyert et al. (2020) applies Yeung's (2018) algorithmic regulation taxonomy to the Uber, Inc. case study and uncovers three essential components of algorithmic regulation—representation of users as data points (as opposed to information gathering using traditional statistics for big data measurements), direction (in place of standard-setting), and intervention (in place of behavior modification). Eyert et al.'s (2020) framework evaluates a broad range of challenges arising from weak governance and unclear industry standards such as tactics for gathering, setting standards, targets or goals (for Uber drivers), and ways of enforcing those standards. Relatedly, Festic (2020) empirically demonstrates the importance of governing online applications that build on algorithmic selection to affect four areas of our everyday life (information, recreation, commercial transactions, and socializing) as well as the data collection, analysis, and sharing activities by AI applications. The author highlights a crucial empirical gap in this area about evaluating users' preferences for governing algorithmic decisions in day-to-day applications used for activities such as loan applications, shopping, or audio-video content recommendations.

Finally, several studies discuss the human-in-the-loop approach to governance of AI, which conveys the argument that algorithmic decision-making cannot entirely replace human judgment in an autonomous application. Scholars advocating this approach argue for human beings' presence and substantive role in any activity involving algorithmic adoption. They do so because human judgment and intelligence are imperative for algorithmic decision-making systems to function effectively, primarily to provide transparency and accountability (Firlej & Taeihagh, 2020; Metcalf et al., 2019). For instance, Binns (2020) points to the conundrum that individual justice cannot be achieved by simply placing human oversight over the algorithmic process. To illustrate, the author points out that the act of determining which exact cases in legal automation will not require a human judge's intervention (as opposed to algorithmic sentencing) will in itself require human intervention and judgment. These arguments point directly toward establishing better occupational norms and forming a collaborative team comprising of human and AI expertise.

Theme: Big Data Collection and Manipulation

Research in this theme recognizes data as an essential *input* to AI and emphasizes societal challenges arising from extensive data collection and manipulation (Alaimo & Kallinikos, 2020). We find that scholars in this area of AI literature have posed two main questions that seek a governance solution: how user data are collected and used (privacy and surveillance) and how transparent the ML algorithms that use these data (bias and opacity) are. These scholars have focused on the challenges generated by data capitalism leading to power imbalances in the platform ecosystem, increased risks to user privacy, and the bias and opacity in AI.

Data capitalism and privacy risk. Big Tech (a name given to the current most dominant IT companies such as Amazon, Apple, Facebook, Google, and Microsoft) is fast becoming a gateway to our daily influx of information (West, 2019; Whelan, 2019). By offering multifarious services and acquiring competitors at an increasing rate, Big Tech platforms are transcending the boundaries of a firm to become an entire ecosystem by themselves. This raises anti-competitive concerns (Romm et al., 2020) and prompts the recalibration of the governance of Big tech's technical and economic infrastructure (Vergne, 2020; Zwitter, 2014). Data capitalism—the capacity of a select few individuals or firms to collect, store, and process large amounts of user-generated data—poses threats to other stakeholders in society (West, 2019). For instance, Rietveld et al. (2020) show that growing platform dominance is problematic for other players in the industry. They uncover through their case study that firms partnering with (giant) platforms, labeled as complementors, find it increasingly difficult to capture value as the partner platform grows. Furthermore, they state that as "a platform becomes increasingly dominant, the platform sponsor's governance strategies shift from being largely supportive of the wider complement population to becoming more selective and geared toward end users" (p. 488). Thus, AI-based firms, particularly, digital platforms, can use the power of big data and user base to imbalance the "value capture" grid in their respective ecosystems and behave opportunistically.

Privacy risk focuses on possible adverse effects of loss of personal information. Data-based firms are prone to collect and process personal data without consent (Yeung, 2018) or with the illusion of informed consent (with data practices buried in fine print). Consequently, Bolin and Andersson Schwarz (2015) highlight that, while it is certainly possible for a concerned user to abandon the AI platform or web service that tracks them, the social costs of staying outside networking platforms are real. These challenges can be explained by an underlying tension between firms' goals toward value creation and individuals' right to privacy.

Bias and opacity in AI. Research in this stream suggests that once data are collected, humans impart meaning and agency to data through two principal means: human programmers (through code) and users (through ML on user-generated data). Scholars have discussed two potential governance risks emanating from data and ML algorithms—bias and opacity. First, we discuss bias in AI. AI applications rely on managers, programmers, and designers to choose appropriate data and code the suitable variables and goals of the algorithm as per the chosen model. Research across computer science, ethics, and management literatures shows that AI can be biased (unfairly prejudiced for or against someone or something) owing to these choices. Choudhury et al. (2020, p. 5) highlight three primary sources of bias in AI: (a) biased training data—arising from bias in the sample selection phase, (b) algorithmic bias—arising from bias in the model designing phase, and (c) input incompleteness—when all relevant information required for search and prediction is not provided. The examples abound. Consider the case of Amazon's gender-biased hiring algorithm as an illustration of algorithmic bias. The algorithms employed in the AI application for recruitment by Amazon were trained on a historical dataset that contained a disproportionately high number of male candidates (Dastin, 2018). As such, the algorithm favored male candidates over females with the same credentials, amplifying the gender divide in an already male-dominated tech industry.

Similarly, Google Photo's racially biased auto-tagging tool, mainly trained on datasets with White faces, gained unpopularity for misclassifying people of color (Hosanagar, 2020). Research has also highlighted that the sensors used in self-driving cars can better detect lighter skin tones than darker ones because the training algorithms use predominantly White subjects' pictures (Haenlein & Kaplan, 2019). As a last example in the legal arena, studies note that decision-support systems used by judges in sentencing can be racially biased because they are predicting the risk of recidivism based on past rulings (Angwin et al., 2016). The governance challenge is putting systems in place to eradicate or minimize any bias such that AI outcomes are fair.

Next, we discuss opacity in AI. Specific subsets of AI, such as neural networks and deep learning algorithms, can be highly non-transparent to the programmer, let alone the average end-user (Buhmann et al., 2020; Burrell, 2016). Such complete or partial lack of interpretability and transparency is called algorithmic opacity and is another area of potential societal concern. While recent ongoing efforts by scientists seek to mitigate programmer's opacity, the options for users of AI are relatively limited (Kane et al., 2021). To illustrate, ML classification tasks such as auto-acceptance of a loan or pattern recognition applications, such as facial recognition, can be a "black-box" for users who do not understand the features used by an AI model to arrive at a decision.

Furthermore, complex AI applications (such as self-driving cars, Cardiograms, or AlphaGo) employ neural networks consisting of deep learning algorithms (Hornik et al., 1989). The inner workings of deep learning algorithms may not always be easy to understand by developers, let alone non-expert users. Thus, for a layperson with little knowledge of data science, being at the receiving end of completely opaque algorithmic decision-making from AI-based devices is bound to be a concern. In addition to algorithmic complexity, Burrell (2016) suggests that two other sources may contribute to AI opacity—technical illiteracy on the part of the human and intentional corporate secrecy.

In terms of governance mechanisms to mitigate the risk of bias and increase transparency, legal regulations lag behind the fast-evolving AI systems. An example was the lethal consequence of an inscrutable algorithmic decision when Uber's self-driving car killed a passenger in Arizona that it did not recognize. Similarly, Kaplan and Haenlein (2019) discuss another incident when Tesla's driverless car running on autopilot mode confused a white truck with a cloud in the sky. Although these incidents brought semi-autonomous AI-based driving systems under new regulatory scrutiny, the companies did not face any criminal charges as the judge found "no basis for criminal liability" for the corporation (Shepardson & Somerville, 2019). Martin (2019) argues that such events underscore the vital need for companies to be accountable for the algorithms they create or apply.

To summarize, research within this theme acknowledges that firms collect data from users with a specific value-creating goal (often profit), and these data are processed through algorithms (usually ML) that are human developed. Therefore, a governance objective highlighted by this body of research is to develop mechanisms and tools that safeguard all members of society, protecting them from privacy breaches and potential societal biases that arise from AI.

Theme: Human–AI Interaction and Collaboration

In this section, we synthesize research that discusses factors influencing processes of human–AI interaction and human–AI collaboration. We find that scholars in this area of the AI literature have focused on factors that influence human trust in AI and the changing nature of human work owing to AI adoption. From a governance perspective, distrust and power asymmetry between organizational players arise when organizations allocate too much agency to an algorithm. Moreover, organizational control is identified as a critical element of governance in that it authorizes and legitimizes how decisions are made and resources are allocated.

Human trust in AI. Research highlights three key features of human trust in algorithms: (a) humans have different standards for measuring algorithmic and human work; (b) human trust in algorithms is subjective and can depend on human characteristics as much as on AI's characteristics and intelligence; and (c) transparent and scrutable algorithms can increase human trust. Through their review of empirical research on the topic of human trust in AI, Glikson and Woolley (2020) advance our understanding of how the human brain perceives robotic AI with a low level of trust, with the level of distrust growing over time. Drawing on factors behind human judgment and trust, they conclude that the level of machine intelligence is a dominant factor in developing human trust in machines. Other factors that play a role in this regard are reliability, transparency, and characteristics of the task performed by the machine.

Relatedly, in his experimental work, Jago (2019) uncovers that audiences have different standards for measuring the authenticity of algorithmic and human work. Study participants find algorithmic work less authentic, even though the output matched their general expectations in multiple work categories of music, food recipes, and so on. In addition, human-guided algorithmic work was deemed more genuine than pure algorithmic work, which underscores the significance of human–AI collaboration and validates wider acceptance of the human-in-the-loop perspective.

Such disparate measurement standards for human and algorithmic work were also observed in experiments by Efendić et al. (2020) conducted to uncover the relationship between perceived effort and prediction quality by humans and algorithms. The authors find that humans judge algorithmic work harshly and are less willing to rely on algorithms when they produce slow predictive results. In contrast, slow response times by humans are considered indicative of a sincere commitment to completing the task, making it more trustworthy. Neighboring this area of research, Logg et al. (2019) study the human characteristics that engender trust in algorithmic work. They demonstrate, with experimental research, that an average person adheres more to algorithmic advice than human advice across a range of subjects. Experts, however, are more likely to discount algorithmic advisors even at the cost of their performance, especially when algorithmic advice was opposed to their own judgment. These studies highlight the variability in perceiving and accepting algorithmic work.

In an interesting departure from the notion of distrust toward algorithms, Raveendhran and Fast (2021) show that reducing human oversight in employee tracking can, in fact, harbor trust and adoption because technology-based tracking reduces participant concern about potential negative judgment. Paying attention to causal mechanisms of human trust/mistrust in algorithms can help decipher the settings and tasks most suited to the widespread acceptance of algorithms. For example, via experimental designs, Lee (2018) illustrates how our knowledge of the source of a decision—algorithmic *or* human—can make a decision appear trustworthy, keeping the nature of the task and skills involved constant. This finding elucidates that if we are not careful in designing AI applications with a human-centric design approach, we risk AI-aversion similar to algorithm aversion. Algorithmic aversion is the human refusal to rely on a superior but imperfect algorithm (Dietvorst et al., 2015). As a follow-up to their work, Dietvorst et al. (2018) recommend providing user control (even if a slight amount) to increase human trust in algorithms. Given the sometimes mixed evidence, it appears that the links between governance and trust are often quite blurry.

AI in the workplace. Current management research indicates that AI can alter the overall workforce structure and reshape organizational control dramatically.

In the last two decades, workplaces have transitioned from employing automated reactive systems such as substitutes for human-powered call centers or loan approvals toward adopting autonomous proactive systems such as auto-fraud detection or image processing for cancer detection (Rao & Verweij, 2017; Tschang & Almirall, 2020). Scholars agree with the notion that jobs are thriving in the age of AI, albeit they are changing (Fleming, 2019; Kellogg et al., 2020). They suggest growth in job opportunities with the advent of AI, calling attention to increasing novel occupations such as algorithmic auditors, curators, and brokers. Tschang and Almirall (2020) emphasize a decrease in routinized middle-skilled occupations and predict proliferation of two kinds of future occupations—the non-routine low-skilled jobs that cannot be automated and the highly skilled technical jobs (albeit in smaller teams), further widening the wage gaps brought by automation in the 1900s. Relatedly, Huang et al. (2019) foresee increasing importance of "feeling tasks" which they claim machines cannot perform better than humans, such as influencing, assisting, and caring for others and their emotions, or resolving interpersonal conflicts and building relationships. Taken together, these studies demonstrate a changing workplace, which may alter our existing lenses of theorizing about organizational control, coordination, and dynamics as we know them.

Kellogg et al. (2020) argue that algorithms are often created and implemented based on the vested interests of powerful organizational actors and claim that algorithms are not neutral tools but "contested instruments of control that carry specific ideological preferences" (p. 383). An example of this power "contest" among organizational players (both humans and AI) is provided by Curchod et al. (2019) in their study of eBay where they examine algorithmic intermediation and employee monitoring in online settings. They identify power asymmetries generated by algorithmic evaluation in online work settings as the "traditional distinction between managers and employees is blurred, and managerial observation is replaced by algorithmic forms of monitoring" (Curchod et al., 2019, p. 2). They illustrate these power asymmetries with the example of online customer evaluations where actors are not physically collocated, and thereby formal appraisals do not occur in traditional settings. However, sellers are made accountable to a diversity of buyers as well as to the platform owner. From a governance perspective, this leads to an imbalanced distribution in favor of platforms that design and implement rewards and sanctions for online sellers. As such, scholars conclude that algorithms are liable to create power imbalances in workplaces owing to inherent algorithmic agency and the agentic actions of those who control these algorithms.

Theme: Organizational Impact of AI Adoption

In this section, we synthesize research that focuses on the outcomes of AI adoption in organizations. Scholars in this area of AI literature have broadly focused on two themes—the organizational outcomes of algorithmic decision-making and organizational creation of economic value with AI. In the current generation of technology, AI does not possess the requisite self-consciousness and self-motivation to resolve ethical conflicts (Braga & Logan, 2017; Munoko et al., 2020). From a governance perspective, this limitation of Narrow AI represents a governance risk when businesses and society rely on it without protective measures. Other governance challenges discussed include algorithmic agency and imbalanced value distribution among stakeholders.

Organizational adoption of algorithmic decision-making. Current research has identified several governance challenges in moving from human to algorithmic decision-making within organizations. Three key features that distinguish the process of human versus algorithmic decision-making are the speed–accuracy trade-off which is comparatively minimal in AI-based decision-making (Shrestha et al., 2019), a loss of outcome interpretability and replicability from human to AI (Haenlein & Kaplan, 2019), and the limited applicability of algorithmic decisions in the current era of weak AI (Raisch & Karkowsli, 2020). It has been argued that these differences make humans and algorithms complementary decision-makers, rather than substitutes (Metcalf et al., 2019). In an automated talent acquisition process, Raisch and Krakowski (2020) highlight that machines cannot fully capture ambiguous predictors and complex variables, such as cultural fit or interpersonal relations. For successful end-to-end operationalization, AI-based technologies still require human supervision because they are limited to the specific task for which they have been trained (Kaplan & Haenlein, 2019; Raisch & Krawoski, 2020). A critical organizational challenge then is to find the appropriate task-distribution balance and its associated governance framework for augmented decision-making involving both humans and algorithms in society.

Research identifies two main challenges pertaining to task-distribution: determining the extent of AI involvement and the effects of algorithmic agency in the workplace. First, there are limitations and boundaries to organizational use of algorithmic decision-making in the current generation of AI. Organizational actors attempting to cross-over these boundaries risk jeopardizing organizational performance, institutional trust, and employee motivation. For instance, in the context of human resources management (HRM), Tambe et al. (2019) explore four challenges that HRM functions may pose for AI adoption (i.e., the complexity of HRM phenomena such as recruitment, training, retention): small data (insufficient data for predictive accuracy of ML algorithms), ethical and legal constraints (e.g., fairness and privacy risk to employees), and negative employee reactions to algorithmic management.

Second, despite the strong narrative that automated decisions are completely agnostic and eliminate human biases, there is a growing body of evidence to prove the contrary. Contemporary research recommends that organizations exercise caution in interpreting and implementing algorithmic decision-making because algorithms can bear intentionality and bias. Algorithms are often created and implemented with specific goals predesigned by powerful actors (Kellogg et al., 2020), which may exacerbate agency problems and power imbalances in organizations. As emphasized in the preceding theme relating to bias and opacity in AI, research in this theme also shows that biases are amplified when ML algorithms train on real-world data potentially contaminated with implicit human bias (Balasubramanian et al., 2020; Choudhury et al., 2020). Moreover, organizations that adopt ML to substitute human decision-making may suffer from sub-optimal results and learning myopia if they rely solely on algorithmic decisions (Balasubramanian et al., 2020; Blohm et al., 2020; D. T. Newman et al., 2020). In sum, research claims that AI is neither interest-neutral nor power-neutral but instead can either carry embedded agency from its creators (or sponsors) or be controlled by powerful agents in the workplace. These relationships need to be structured within an agreed-upon governance system.

Creating economic value with AI. Contemporary research in this stream focuses on value creation, capture, and sustained competitive advantage through AI. Research suggests that with growing capacity to collect, process, and store data, businesses are likely to follow a trajectory of acquiring increasingly large data resources—both to improve their offerings and to gain competitive advantage with an overall shift from "exclusivity in technology to exclusivity in data" (Hartmann & Henkel, 2020, p. 359). As a result, some scholars have claimed that the time is ripe for firms to develop new business models (Clough & Wu, 2020; Gregory et al., 2020), highlighting that value creation in AI-based digital platforms relies heavily on network effects. By virtue of such network effects, user data (when processed using ML algorithms) help firms to build more efficient and innovative solutions for their customers (and themselves), which in turn attracts more customers to the platform, thereby bringing further business opportunities. Gregory et al. (2020) label this new variation of a firm's reliance on networks as "data network effects" and theorize that data volume from AI-based platforms is a key indicator of perceived value for users.

Unlike other resources, however, more data are not always a primary source of competitive advantage. For example, Microsoft's chatbot Tay, which was launched on Twitter in 2016 after having tasted success in a similar AI-based service in China (Xiaoice) had to shut down within hours of its release, bringing significant reputational and economic harm. The problem was not the chatbot's unpopularity or lack of engagement. If anything, it was the opposite; with an abundance of "racist, fascist and sexist tweets," Tay quickly learned to imitate this behavior in its interactions (Hosanagar, 2020, p. 4) with new users. Such incidents are indicative of a need for developing effective governance mechanisms alongside building effective business models.

Findings and future research suggestions from this stream of research point toward governance-based sustainability, that is, the most valuable and sustainable business models will be built on a set of effective governance frameworks (Flyverbom et al., 2019; Rietveld et al., 2020; Vergne, 2020). This includes well-governed data practices that respect user privacy with an equitable sharing of appropriated value with stakeholders. In sum, we know that AI will create value for business and society, but there are many other open questions such as value for whom? will it also destroy value for others? and are we, as a society, comfortable with the answers. Governance solutions to some of these questions may come with further economic challenges of supplementing the "algorithmic gaze" (Newlands, 2021)? For instance, organizations with augmented teams, where machines and humans are integrated to complement each other, will explore adaptations to their business models to accommodate additional costs of keeping a human-in-the-loop (Newlands, 2021; Raisch & Krawoski, 2020).

A Framework for Governance of AI—Governance Modalities

Beginning in the mid-1990s, Lawrence Lessig developed a socio-economic view on the regulation of cyberspace and argued that internet users are influenced via constraints on their actions through four modalities: the law, the market, social norms, and the architecture of the network. We extend Lessig's (1998) New Chicago School theory to the governance of AI and propose that governance of AI occurs at the intersection of these four modalities because each of these modalities regulates the AI ecosystem by imposing a set of constraints on either the process of data collection, algorithmic manipulation,

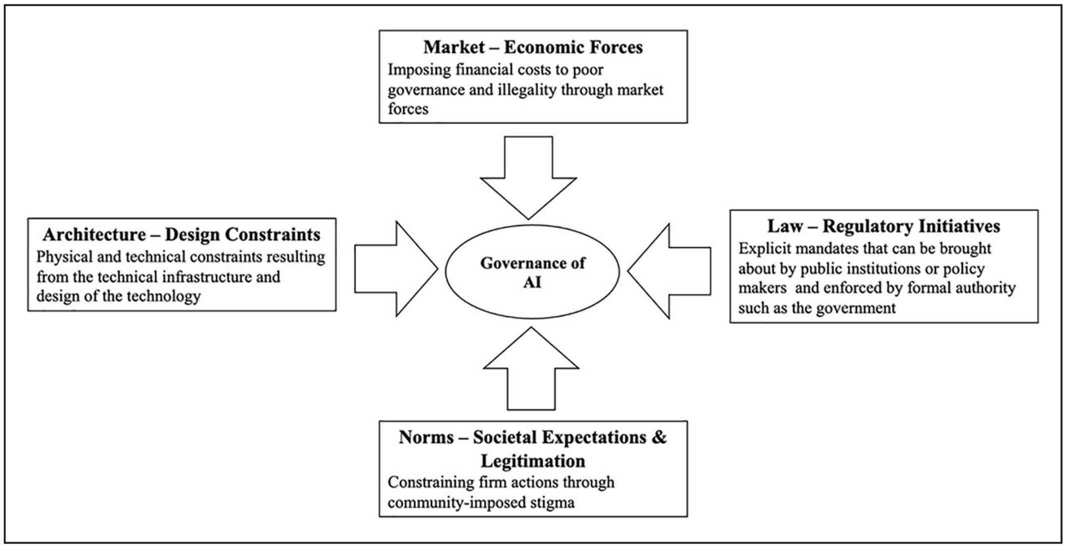

Figure 4. Governance modalities.
Note. AI = artificial intelligence.

or use of the AI application. We draw on this framework to examine where governance challenges of AI reside and from what modality angle can best be tackled. We describe each modality in greater detail and discuss how this governance framework holds relevance for a technology as disruptive today as the internet was four decades ago. We illustrate each governance modality in Figure 4. We close by discussing how the three research themes of AI are interconnected through these governance modalities.

First, the governance modality of *law* includes explicit mandates that can be brought about by public institutions or policy makers and enforced by a formal authority such as the government. This governance modality applies relatively straightforward to AI. For example, emerging challenges with data privacy have led to regulatory initiatives to govern the collection and use of data (i.e., the California Consumer Privacy Act [CCPA] and General Data Protection Regulation [GDPR]), introducing a renewed focus on privacy among firms. Although firms have consequently updated their privacy policies, terms of use, and user agreements, Andrew and Baker (2019) argue that the privacy laws under GDPR do not offer sufficient individual protection to users against the risks associated with the collection, analysis, and trade of their data. Li and Sarkar (2014) claim that one of the reasons these laws fall short is the practice of personal data being shared with a third party without user consent or control. They point to technical loopholes in the existing legal framework for events such as regression attacks (made using regression trees to reveal individuals' sensitive data). The authors empirically demonstrate that existing privacy-preserving techniques are inadequate to shield against such attacks, and in fact, some of the current norms could even make it easier for attackers to access sensitive data. As such, privacy risk is still an ongoing governance challenge, but the law is slowly catching up.

Second, the governance modality of *the market* comprises economic forces, including the forces of supply and demand, that acts as an invisible (governing) hand by enforcing financial costs of poor governance standards. To illustrate in the realm of AI, Agrawal et al. (2018) propose that firms that invest in protecting user privacy should be perceived by users as better-performing firms and hence have higher legitimation and reputation. For example, Apple, Salesforce, Adobe, Uber, Dropbox, among others, have placed a strategic bet by investing heavily in protecting privacy (Agrawal et al., 2018; Benner & Mozur, 2016; Hosanagar, 2020). Thus, market forces of increased demand for privacy-protecting practices can promote platforms to adopt effective governance mechanisms that safeguard user privacy. These governance practices are market-driven and grant legitimation to firms.

ML systems introduce the need for a different type of governance. Legal constraints, which include explicit definitions of illegal activities, cannot prove effective on ML and deep learning algorithms as they identify patterns to make predictions without explicitly revealing the rules being applied. They are designed by the system designer(s) to optimize specific objectives and goals, often overlooking the interests of all stakeholders. As such, norms can more effectively govern opaque processes that rely on predefined features, assigned weights, and typical training data. The *norm* governance modality refers to forces constraining (organizational and individual) actions "through the stigma that community imposes" (Lessig, 2006, p. 124). In simple terms, norms can help establish tacit societal expectations of everyday governance practices. For example, in the realm of AI interpretability, scholars have recommended that striving for transparent functioning when possible (e.g., supervised algorithms) or for maximum interpretability (e.g., in case of neural networks) helps harbor and enhance user trust (Koo & Eesley, 2021). As ML algorithms detect, classify, and predict outcomes, which are often followed by automated actions, these interpretability efforts must go hand-in-hand with organizational accountability of algorithmic decisions. Driven by a societal push to adopt new measures for fair, accountable, and transparent applications of AI, organizations can adopt new norms and procedures such as auditing AI algorithms (Shrestha et al., 2019). Even as the process of auditing algorithms can be fraught with IPR (Intellectual Property Rights) concerns and thereby not enforced by law, firms can seek digital auditing for their data-handling practices because it can help facilitate trust across organizational and societal actors. To illustrate, recent empirical work by Schafheitle and colleagues (2020) examining how AI adoption modifies trust among various organizational actors finds that accountable and transparent practices of data collection and algorithmic tracking (e.g., smart ID badges, wearable GPS devices, or smart toilets) generate higher employee trust, leading to an increase in adoption of AI practices in the workplace.

Finally, the physical and technical constraints resulting from the infrastructure and design of technology fall under the *architectural* governance modality. Such AI-design approaches include human-in-the-loop control (Firlej & Taeihagh, 2020), which calls for substantive human involvement alongside algorithms for effective and accountable AI systems. As discussed in the section on existing research discussing the governance of AI, scholars show a broader social acceptance of AI applications that are human-centric in their approach. AI-applications in line with this approach are built with a culture of keeping human needs first (and business needs second), where the user is always actively engaged with a system with the ability to override or intervene. Architectural governance also includes non-technical design implementation. For example, at the level of the platform, Rietveld and colleagues (2020) recommend three types of governance changes that enable complementors to create and capture value. These governance changes are mostly targeted at digital platform ecosystems, where the platform-owning organization can update a platform's technological infrastructure to reduce the entry barriers for complementors and facilitating more choice of complements in the platform's ecosystem. In sum, the architecture governance modality draws on the nature of the infrastructure to dictate the rules of transparency and accountability.

Although effective by themselves, these four modalities govern AI much better in tandem. To illustrate, Gregory and colleagues (2020) urge business leaders to employ governance norms of a firm's data assets, embracing user-friendly products and services, thereby creating greater value for platform users. Table 3 exemplifies some of the intersections among the four governance modalities and the different AI challenges. The main takeaway is that we find an alignment across the governance modes, that is, among the regulation developed by policy makers and approved by legislators, the market rules defining the cost of non-complying with the market forces, the societal norms legitimizing the laws and the market forces, and the appropriate way of doing things, and ultimately the technological and infrastructure rules defining how AI is governed.

Even as some governance modalities might be more salient or dormant than others, these four modalities are all interdependent. For example, if we take one of the rows in Table 3, the challenge of AI opacity requires new rules at the infrastructure level (architecture), but it also carries financial

Table 3. Review Themes in a Governance of AI Framework.

		Governance modalities			
Review themes	Societal challenges	Law	Market	Norms	Architecture
Inputs	Big data collection and manipulation				
Data capitalism and privacy risk	*Power imbalance*	Anti-competitive laws	User boycott of unethical practices		Structural, boundary-spanning and redistributive measures to alter technological infrastructure
	Loss of privacy	Privacy protecting legal framework	User aversion toward weakly governed AI systems prone to breaches	Level of societal expectation and tolerance for privacy	Adopting privacy-by-design for users
Bias and opacity in AI	*Algorithmic bias*	Auditing of data-handling practices		Augmented (human–AI) teams	Disclosures of technical processes
	Opacity of internal processing	Digital auditing (including model)	Customer preference for maximum interpretability	Augmented (human–AI) teams	Accessible blueprint of ML model

(continued)

Table 3. (continued)

		Governance modalities			
Review themes	Societal challenges	Law	Market	Norms	Architecture
Process	Human–AI interaction and collaboration				
Human trust in AI	*AI-aversion*		User feedback	Human-centric AI system	Increasing user-control
AI in the workplace	*Reinforced power asymmetries and algorithmic agency*	Industry standards to reduce labor market vulnerabilities	Incorporating employee feedback	Participatory design incorporating feedback from stakeholders	Aggregated decision-making or human-in-the loop
Outputs	Organizational impact of AI adoption				
Organizational adoption of algorithmic decision-making	*New occupational responsibilities*			Augmented (human–AI) teams	Distinct task-distributions (given limitations of AI)
Creating economic value with AI	*Imbalanced property rights division*	Anti-competitive laws		Setting industry standards for platform complementors	

Note. AI = artificial intelligence.

costs for not being more transparent (market) as well as societal disapproval (norm), while the law might not be such an effective governance modality because it is tough to enforce. Even though we have discussed the challenges in AI within the three research themes, Table 3 shows that when examining their governance, they are all intertwined not only in terms of governance modalities, but also in terms of AI challenges. For example, if we think about the columns in Table 3, it is clear that a given hard law to improve data collection practices will have an effect on improving user trust in how those data are used. Similarly, societal norms affecting the different AI challenges have to be congruent. Hence, if there is an effort to demand augmented or human-centric AI teams normatively, this will spill over into having greater trust in AI.

There exist some illustrations in the literature examining options in governance modalities. For example, Arthur and Owen (2019) contrast statutory (law) and self-imposed governance (architecture) as two approaches to governing data practices with the overall aim of competitive advantage to the firm. They analyze big data's acquisition, manipulation, and commodification, in the banking and retail sectors and highlight the competitive advantages of developing trust through governance. Their case study suggests that firms across all sectors are likely to gain a relative advantage if they can develop ethical behavior and good governance practices. Thus, user-behavior and purchasing patterns can increasingly incentivize firms to strategically invest in building trustworthy AI and data practices.

Future Research

The study of AI from a governance perspective presents tremendous opportunities given the nascent stage of the field and the significant implications for societal advancement and alleviation of grand societal challenges. This section is divided into two parts. First, we highlight five governance paradoxes that emerged from our review of AI literature, where future scholars of AI could devote more attention. Second, we offer recommendations on how theory-building and extending existing management theories can shed light on novel theoretical questions emerging from sustainable adoption and effective governance of AI. We conclude by making a call for a unifying theory of governance of AI.

Balancing Societal Trade-Offs Through Governance

One of the conclusions that emerged from our literature review is the conceptualization of governance as a trade-off concept with respect to other societal goals—in that more of one dimension directly limits another dimension because they are paradoxical. We discuss five governance trade-offs and identify research gaps in the governance of AI adoption. While the empirical reality is often far from being dichotomous, we present these ideal type trade-offs as binary choices because it enables us to move in the direction of asking complex and consequential questions about the effectiveness of current AI governance.

Governance versus innovation. The development of information and communication technology (ICT), closely linked to the emergence of AI, has gone hand-in-hand with deregulation and neoliberalism (Lundvall, 2017). The extreme ends of a continuum in the governance spectrum are hard (government) enforced regulatory governance versus laissez-faire innovation (self-governance), allowing for more creativity. A trade-off that surfaces is whether our society needs more regulation and data protection, or instead an incentivized open digital economy to achieve better innovation output and economic growth? We encourage scholars to analyze the repercussions of choosing more regulation over an open digital economy for all societal actors involved, including organizational innovators, regulatory authorities, and the general public. Policy makers and businesses will benefit greatly from determining the extent of regulation needed for algorithmic activities before designing processes by which AI can be governed (Festic, 2020; Khanna, 2018). Future research can explore the consequences of leaning on the side of regulation and critically analyze potential solutions such as innovation sandboxes where AI innovators and regulators set up formal procedures around the use of innovation, reduce uncertainty, and raise transparency for all stakeholders.

Reforming versus strengthening a surveillant state. A difficult balance in an increasingly digital world is between the need for national protection and respecting citizens' right to privacy (Redden, 2018). Essentially, there is no clear boundary where governmental protection ends, and an Orwellian surveillance state begins. Recent examples include China's use of facial recognition technology (Colback, 2020) and authoritarian repression of Vietnamese protestors facilitated by Facebook and YouTube (Ratcliffe, 2020). Future governance scholars can explore possible solutions such as deliberative governance (Donahue & Zeckhauser, 2012; Scherer & Voegtlin, 2020), which involves bringing public and private stakeholders together in a collaborative process of governance. Such measures can build trust in the government decision-making and legislation around surveillance.

Distributed versus concentrated power. A central challenge to AI adoption that we highlight in our review is that societal actors, including but not limited to private firms, are asked to undertake new business responsibilities arising from changing social realities, new forms of capitalism, evolving business models, and emergent institutional forces. However, the set of organizational and societal actors involved in the AI ecosystem do not necessarily possess complete knowledge of the entire value chain that AI presents. As each actor (users, coders, managers, board members, policy makers) may only contribute partially to effective governance mechanism, there is a high chance of "normal accidents" that are waiting to happen (Nunan & Domenico, 2017), exacerbated by a lack of accountability for such accidents. Fortunately, practitioners and scholars have shown a growing interest in the field of Technical Social Responsibility (TSR) or Digital Social Responsibility (DSR). Assuming TSR as a governance practice would mean "a conscious alignment between short-and medium-term business goals and longer-term societal ones" (Bughin & Hazan, 2019, p. 2). Current literature also does not discuss the classification and segregation of tasks necessary to effectively distribute accountability and responsibility between human–AI ensembles. There also is a need to establish a task-balance of optimizing workload between the two. This is a research gap that future studies can address.

Algorithmic efficiency versus fair data practices. Scholars have pointed to the fundamental privacy-efficiency trade-off in decisions made via ML algorithms (Choudhury et al., 2020; Hosanagar, 2020; Kane et al., 2021). To achieve high accuracy from predictive and prescriptive ML algorithms, programmers must train them on data that mirror reality. While this is optimal for the algorithm's functionality, research suggests that an organization has to cope with data biasedness, user privacy concerns, and regulatory limitations by optimizing the level of data anonymization at the cost of algorithmic efficiency (Y. Chen et al., 2021). A conceptual relationship that is not yet addressed is the trade-off between transparency and efficiency of AI models (Lambrecht & Tucker, 2019). We encourage future scholars to conduct in-depth case studies to analyze contemporary organizations with explainable and trustworthy AI governance practices, so as to explore processes and outcomes of making AI systems sufficiently transparent without losing significant efficiency. For example, it would be fruitful to analyze whether investing in higher "explicability" or transparency of an AI application to foster increased consumer trust would lead to higher financial performance than an opaque high-performing AI application. We believe that it is essential for future studies to examine the user demand for such transparency in AI applications to address the need and strength of governance forces (if at all).

Biased AI versus societal bias. Another critical but challenging area that the literature points us to is the governing and regulating training data to prevent amplifying human biases. Interested scholars can investigate governance measures that support technical decisions, such as, what is the relationship between the amount of historical training data and AI's biasedness (e.g., curvilinear)? What governance mechanisms could be put in place (architectural and informal practices) to minimize bias and increase transparency? Are there empirical scenarios in which algorithmic decision-making itself is bound by the human programmer's cognitive capacity and the training data that ML models train on? Future research can contrast organizational decision-making before and after the intervention of AI adoption. Thus, this area presents a promising avenue for future scholars to explore successful cases of business viability with a sustainable goal of building a fair and transparent AI.

Theoretical Contributions

In this section, we suggest how we can better understand the governance of AI using three existing theories in management. We discuss ways in which building upon and extending these theories could help form a unifying theory of AI's effective governance, addressing the challenges that emerged from extant research and the governance trade-offs just discussed.

Institutional theory. Institutions are closely linked to governance in that they dictate the nature of governance. We discuss three different schools of institutional theory (Aguilera & Grøgaard, 2019). First, we refer to institutions as the social structures (cognitive, normative, and regulative) that provide organization and meaning to social life (Scott, 2013). Scholars of institutional theory have emphasized how institutional processes influence organizations and play a consequential role in legitimizing organizations. In our discussion of the governance modalities, norm and market modalities operate through institutional forces of legitimation and trust. Organizations striving to gain a competitive advantage via effective governance practices are likely to gain legitimation and trust from users, media, and employees. For instance, big tech firms compete for a broader user base partly granted by their data privacy practices. Future research can explore the conceptual validity of how good governance practices in AI are likely to diffuse into the field by social learning or mimetic isomorphism and can arguably be a source of organizational legitimation.

Another conceptualization of institutions is given by North (1990) where they are seen as rules of the game that regulate organizational and individual behavior through formal (laws) and informal (norms) ways. The governance modalities of norm, market, and law that we have discussed are in accordance with this perspective and rely on the external institutional environment for these modalities (hard laws and soft laws or norms) to become enforceable and legitimate. Future research can investigate the effects of institutional variation across regions, countries, and industries on organizational responses.

Finally, future researchers are encouraged to revisit leaders' roles and responsibilities in times of perennial organizational transformation and highlight implications for AI. Scholars can draw upon the role of leadership values and responsibilities (Selznick, 1957) in successful AI adoption in organizations. In our review theme on *Organizational Implications of AI*, we stress the risks of handing over decision-making to algorithms without taking the external environment into account. Based on extant research, we propose that as long as algorithms lack moral imagination, algorithmic decision-making is likely to be an ineffective substitute for organizational leadership. If anything, effectively managing AI applications and solutions is expected to add to a "leadership's responsibilities." Future scholars can build on Selznick's central finding that most of organizational life is an interplay between both technical and institutional forces (Besharov & Khurana, 2015), to analyze organizations using analytical solutions in order to support and aid decision-making. Future scholars can draw on institutional theory and help corroborate claims by recent research (see Metcalf et al., 2019; Murray et al., 2020) that augmented decision-making, as well as a conjoined agency can lead to societal advancement as long as the final decision rests in the hands of humans.

Property rights and stakeholder theory. Property rights are the social institutions that define the range of sanctions, and the range of privileges granted to individuals to specific resources and typically include three bundled rights: the right to use, right to appropriate stream of economic rents from, and the right to change the form or transfer the asset (Libecap, 1989). Based on our literature review, we argue that digital property rights are becoming increasingly contested and undefined—as there is a conceptual tension between the right of the firm that owns a data-driven AI platform/application and the users of such a service. Data lie at the intersection of public and private ownership. Although data can be considered privately generated (by each user), as soon as algorithms on digital platforms process them, they share the non-rivalrous nature of a public good. Appropriation of such digital property rights can be difficult in a weak legal and institutional environment. Consequently, platform owners may reserve the right to deny services to users who are unwilling to part rights with their digital information (ranging from browser cookies, geo-location, and personal demographics in some cases). Future studies can identify when privacy rights over user data supersede the rights of companies who are expending resources to collect, process, store, and utilize our data without charging us for the data-driven services they offer. Scholars may also find it fruitful to extend theoretical underpinnings on the allocation of digital property rights.

In our view, extant research on economic value from AI focuses exclusively on ways of value creation and capture while ignoring questions about sharing this value (Bacq & Aguilera, 2021). Essentially, users create value through their participation on platforms and open-source systems. Platform-owning firms in turn add value via proprietary algorithms before eventually shifting this value in their own favor by substantially appropriating a large part. Thus, AI-based firms can be seen to strategically favor one stakeholder (shareholders) versus others (users) by adopting a creating–capturing–shifting nature of economic value generation by AI. Future empirical work can help provide evidence of this value shift and associated benefits for each stakeholder.

Future scholars can develop new definitions of firm boundaries given that the platform users and complementors lie outside the boundaries of the platform-owing firm while their data and applications, respectively, are centrally controlled by the platform-owning firm. This argument about the boundaries of value governance suggests the need for revising current theories of stakeholder governance (Amis et al., 2020). However, going beyond contemporary research, the question will not be if businesses will create value from AI but rather how they will do so? Future scholars can ask whether organizational actors can potentially exercise greater control over the scale of AI's impact and value distribution through well-planned governance frameworks.

Agency theory. Jensen and Meckling (1976) conceptualize the firm as a nexus of contracting relationships among individuals and argue that the firm is not an individual but is a legal fiction that serves as a focus for a complex process in which the conflicting objectives of individuals are brought into equilibrium within a framework of contractual relations. They define agency costs as the sum of the principal's monitoring expenditures plus the agent's economic bonding expenditures plus the residual losses. Future research can explore the nature of such costs and contracts among the principal (users) and agents (AI-product and service provider).

Agency costs occur when there exists a separation between ownership and control. Under such conditions where the interests of owner(s) and managers often diverge, and managers possess discretionary power, managers are prone to opportunistic behavior. Algorithmic decision-making practices are likely to change the dominant conversations in governance literature (for instance, regarding the agency relationships and incentive alignment). Consider, for example, the actors who partake in the AI ecosystem. Typical actors in this set include employees who design algorithms and innovate applications of AI, managers and board members who act as decision makers in AI firms (establishing industrial norms), shareholders invested in firms, policy makers and legislators who develop and enforce the rules surrounding the use of AI-based applications, and other stakeholders such as users and employees of firms who bear the consequences of AI-adoption. The interests and incentives of each of these actors are likely to differ and may even conflict. For instance, shareholders might be trying to maximize their economic and information incentives and while users, as stakeholders, seek to get good products or services. Scholars can help identify conditions and factors whereby a firm's economic and social interests are aligned with building ethical data and algorithmic practices.

To conclude, we find that there are existing theoretical concepts and mechanisms that can help us further our collective understanding of the governance of AI. However, one of the main insights from this review is that AI requires an eco-system perspective to assess how different parts (of theory and practice) contribute to the whole. The time is ripe for a comprehensive theory of AI governance that identifies the goal of AI-adoption, where the boundaries are, the roles and interests of multiple involved actors, and their underlying governance mechanisms to achieve these goals. To achieve a thorough understanding of this seemingly complex process, we encourage integrating several conceptual perspectives that account for empirical context, actors' rights, and interests at various levels of abstraction (i.e., an individual, an organizational as well as a national level). We envision such research to employ multiple methods in the empirical research and benefit from cross-pollination of research ideas among disciplinary fields. We hope that the discussions in this review foster scholarly interest and provide the groundwork for such theory development.

Recommendations to Businesses and Policy Makers

In alignment with arguments presented throughout this review article, we recommend managers to engage in skill enhancement, create training plans for themselves and the employees, and democratize the process of AI adoption with inputs from stakeholders factored into the process. Based on our incomplete understanding of the complexities involving algorithmic decision-making, organizations may be well served to keep the "H" in HR until we have established frameworks for fair, transparent, and accountable use of AI adoption in organizations. We are yet to build regulatory and governance systems to answer fundamental questions such as who is really accountable when algorithms falter—the programmer? the manager? the top executives? Or maybe all of them. This distributed responsibility amounts to an absence of accountability, for we certainly cannot keep algorithms behind bars.

To nurture responsible business leaders of tomorrow, we recommend that universities educate students on topics at the intersection of data science and society to help them develop critical thinking about the social manifestations of algorithmic power (Kaplan & Haenlein, 2019; Leicht-Deobald et al., 2019). Such a course taken by societal members at large can help them develop awareness and reflective skills about their digital footprint, data privacy, and ownership rights. Today, we have human systems that are augmented by AI rather than vice versa. Existing research seems to advocate for more precise governance practices harnessing this relationship. We entrust policy makers and regulators to be careful in designing governance frameworks for tech firms on society's behalf because we are aggregating individual wills and desires to build collective solutions on behalf of disparate individuals.

We also encourage scholars to apply and advance our proposed governance framework to other organizational settings and technologies. We believe that this framework has wider application (beyond ML and AI) to the governance of other complex technologies such as blockchain, or relatively newer forms of technical infrastructure, such as multi-sided platforms, where governing rules are ambiguous or weakly defined. We have illustrated the role of each governance modality in Figure 4 to provide the reader with a snapshot of how these four modalities act upon in different ways through integration of laws, code, norms and culture, as well as market and institutions to facilitate coordination and determine the governance of a particular technology or hybrid organization, which is AI, in our case.

Conclusion

The AI revolution is introducing significant structural and institutional transformations by challenging the legal environment (e.g., existing laws and regulations), industry practices (e.g., standard settings), emerging organizational forms (e.g., platforms), new forms of labor (e.g., algorithmic auditors), and leadership responsibilities (e.g., shifting authority regimes and new professional roles). Consequently, the governance mind-set that worked for other technologies will likely be insufficient for the governance of AI and its rapid technological advancements. On one hand, AI-based applications have increasing access to massive datasets with growing adoption and rapid innovation. But, on the other hand, we are yet to form effective governance and regulatory practices on AI's fair, equitable, and accountable deployment.

In our literature review, we find that management scholars have largely focused on the organizational impact of AI-adoption, and the firm capability enhancement to create greater economic value from AI, while placing relatively less emphasis on the social costs of doing so and its regulation. Although it is vital to study the organizational impact of AI (the largest theme as per this review), the field of management and strategy is uniquely positioned to support organizations that adopt AI sustainably. In this article, we have discussed how governance may harness AI's power to help solve societal problems without creating or amplifying societal inequalities. Through managerial implications and recommendations, research in our field holds the potential to assist organizational actors in their ongoing journey of AI adoption.

Acknowledgments

We would like to thank our associate editor Jill Brown and the three reviewers for their encouraging and constructive feedback and guidance. We would also like to thank audiences at the Academy of Management 2020 and Global Strategy and Emerging Markets Conference at Cornell University 2020. We thank the Center for Professional Responsibility in Business and Society (Gies College of Business, University of Illinois, Urbana-Champaign) for financial support toward this project. We thank Gretchen Winter, Mark Peecher, and Ruthanne Huising, for their ideas and encouragement toward this project.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Ruth V. Aguilera https://orcid.org/0000-0002-1144-1499

Supplemental Material

Supplemental material for this article is available online.

Notes

1. Academic journals included in the bibliometric searches are as follows: (a) *Academy of Management Annals*, (b) *Academy of Management Discoveries*, (c) *Academy of Management Journal*, (d) *Academy of Management Learning*

and Education, (e) *Academy of Management Perspectives*, (f) *Academy of Management Review*, (g) *Administrative Science Quarterly*, (h) *Business & Society*, (i) *Entrepreneurship Theory and Practice*, (j) *Journal of Business Ethics*, (k) *Journal of International Business Studies*, (l) *Journal of Management*, (m) *Journal of Management Studies*, (n) *Management Science*, (o) *MIS Quarterly*, (p) *Organization Science*, (q) *Organization Studies*, (r) *Organizational Behavior and Human Decision Processes*, (s) *Research Policy*, (t) *Strategic Entrepreneurship Journal*, (u) *Strategic Management Journal*, and (v) *Regulation & Governance*.

2. *California Management Review* (CMR) has published some seminal work on AI, and it was included in the FT (*Financial Times*) journal list in the past.
3. Including these articles in our first step gave us a chance to find governance-related research in books and editorials.

References

Articles preceded by an asterisk are included in Table A in the Online Appendix.

*Adner, R., Puranam, P., & Zhu, F. (2019). What is different about digital strategy? From quantitative to qualitative change. *Strategy Science*, *4*, 253–261. https://doi.org/10.1287/stsc.2019.0099

Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction machines: The simple economics of artificial intelligence*. Harvard Business Press.

Aguilera, R. V., & Grøgaard, B. (2019). The dubious role of institutions in international business. A road forward. *Journal of International Business Studies*, *50*, 20–35.

Aizenberg, E., & van den Hoven, J. (2020). Designing for human rights in AI. *Big Data & Society*, *7*(2), 2053951720949566.

*Alaimo, C., & Kallinikos, J. (2020). Managing by data: Algorithmic categories and organizing. *Organization Studies*, *42*(9), 1385–1407. https://doi.org/10.1177/0170840620934062

Allen, M. (2018). You snooze, you lose: Insurers make the old adage literally true. Pro Publica.

Amis, J., Barney, J., Mahoney, J. T., & Wang, H. (2020). From the editors—Why we need a theory of stakeholder governance—And why this is a hard problem. *Academy of Management Review*, *45*, 499–503.

*Andrew, J., & Baker, M. (2019). The General Data Protection Regulation in the age of surveillance capitalism. *Journal of Business Ethics*, *168*(3), 565–578.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias. *Pro Publica*. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

*Arthur, K. N. A., & Owen, R. (2019). A micro-ethnographic study of big data-based innovation in the financial services sector: Governance, ethics and organisational practices. *Journal of Business Ethics*, *160*, 363–375.

Bacq, S., & Aguilera, R. V. (2022). Stakeholder governance for responsible innovation: A theory of value creation, appropriation, and distribution. *Journal of Management Studies*, *59*(1), 29–60.

*Balasubramanian, N., Ye, Y., & Xu, M. (2020). Substituting human decision-making with machine learning: Implications for organizational learning. *Academy of Management Review*. Advance online publication.

Benner, K., & Mozur, P. (2016, February 20). Apple sees value in its stand to protect security. *The New York Times*. https://www.nytimes.com/2016/02/21/technology/apple-sees-value-in-privacy-vow.html

Besharov, M. L., & Khurana, R. (2015). Leading amidst competing technical and institutional demands: Revisiting Selznick's conception of leadership. In M. S. Kraatz (Ed.), *Institutions and ideals: Philip Selznick's legacy for organizational studies (Research in the sociology of organizations)* (Vol. 44, pp. 53–88). Emerald Group.

*Binns, R. (2020). Human judgment in algorithmic loops: Individual justice and automated decision-making. *Regulation & Governance*, *16*(1), 197–211. https://doi.org/10.1111/rego.12358

*Blohm, I., Antretter, T., Sirén, C., Grichnik, D., & Wincent, J. (2020). It's a peoples game, isn't it?! A comparison between the investment returns of business angels and machine learning algorithms. *Entrepreneurship Theory and Practice*, 1042258720945206.

*Bolander, T. (2019). What do we lose when machines take the decisions? *Journal of Management and Governance*, *23*, 849–867. https://doi.org/10.1007/s10997-019-09493-x

*Bolin, G., & Andersson Schwarz, J. (2015). Heuristics of the algorithm: Big Data, user interpretation and institutional translation. *Big Data & Society*, *2*(2), 2053951715608406.

Braga, A., & Logan, R. K. (2017). The emperor of strong AI has no clothes: Limits to artificial intelligence. *Information*, *8*(4), Article 156.

*Brock, J. K. U., & Von Wangenheim, F. (2019). Demystifying AI: What digital transformation leaders can teach you about realistic artificial intelligence. *California Management Review*, *61*(4), 110–134.

*Brynjolfsson, E., Hui, X., & Liu, M. (2019). Does machine translation affect international trade? Evidence from a large digital platform. *Management Science*, *65*(12), 5449–5460.

Brynjolfsson, E., & Mcafee, A. (2017). The business of artificial intelligence. *Harvard Business Review*, *7*, 3–11.

Bughin, J., & Hazan, E. (2019, August). Can artificial intelligence help society as much as it helps business? *The McKinsey Quarterly*.

*Buhmann, A., Paßmann, J., & Fieseler, C. (2020). Managing algorithmic accountability: Balancing reputational concerns, engagement strategies, and the potential of rational discourse. *Journal of Business Ethics*, *163*(2), 265–280.

*Burrell, J. (2016). How the machine "thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1), 2053951715622512.

*Cennamo, C., Marchesi, C., & Meyer, T. (2020). Two sides of the same coin? Decentralized versus proprietary blockchains and the performance of digital currencies. *Academy of Management Discoveries*, *6*, 382–405.

*Chalmers, D., MacKenzie, N. G., & Carter, S. (2020). Artificial intelligence and entrepreneurship: Implications for venture creation in the fourth industrial revolution. *Entrepreneurship Theory and Practice*, *45*(5), 1028–1053.

*Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, *36*, 1165–1188.

*Chen, Y., Pereira, I., & Patel, P. C. (2021). Decentralized governance of digital platforms. *Journal of Management*, *47*(5), 1305–1337.

*Chenou, J. M., & Radu, R. (2019). The "right to be forgotten": Negotiating public and private ordering in the European Union. *Business & Society*, *58*, 74–102.

*Choudhury, P., Starr, E., & Agarwal, R. (2020). Machine learning and human capital complementarities: Experimental evidence on bias mitigation. *Strategic Management Journal*, *41*(8), 1381–1411.

*Christin, A. (2017). Algorithms in practice: Comparing web journalism and criminal justice. *Big Data & Society*, *4*(2), 2053951717718855.

*Clough, D. R., & Wu, A. (2020). Artificial intelligence, data-driven learning, and the decentralized structure of platform ecosystems. *Academy of Management Review*, *47*, 184–189.

Colback, L. (2020, October 15). The impact of AI on business and society. *Financial Times*. https://www.ft.com/content/e082b01d-fbd6-4ea5-a0d2-05bc5ad7176c

*Curchod, C., Patriotta, G., Cohen, L., & Neysen, N. (2019). Working for an algorithm: Power asymmetries and agency in online work settings. *Administrative Science Quarterly*, *65*(3), 644–676.

Dastin, J. (2018). *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters.

*Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology*, *144*, 114–126.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, *64*, 1155–1170.

Donahue, J. D., & Zeckhauser, R. (2012). *Collaborative governance: Private roles for public goals in turbulent times*. Princeton University Press.

*Dourish, P. (2016). Algorithms and their others: Algorithmic culture in context. *Big Data & Society*, *3*(2), 2053951716665128.

*Efendić, E., Van de Calseyde, P. P., & Evans, A. M. (2020). Slow response times undermine trust in algorithmic (but not human) predictions. *Organizational Behavior and Human Decision Processes*, *157*, 103–114.

Elkin-Koren, N. (2020). Contesting algorithms: Restoring the public interest in content filtering by artificial intelligence. *Big Data & Society*, *7*(2), 2053951720932296.

*Etzioni, A. (2018). Apple: Good business, poor citizen? *Journal of Business Ethics*, *151*, 1–11.

*Eyert, F., Irgmaier, F., & Ulbricht, L. (2020). Extending the framework of algorithmic regulation. The Uber case. *Regulation & Governance*, *16*(1), 23–44. https://doi.org/10.1111/rego.12371

*Felten, E., Raj, M., & Seamans, R. (2021). Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses. *Strategic Management Journal*, *42*(12), 2195–2217.

*Festic, N. (2020). Same, same, but different! Qualitative evidence on how algorithmic selection applications govern different life domains. *Regulation & Governance*, *16*, 85–101. https://doi.org/10.1111/rego.12333

*Firlej, M., & Taeihagh, A. (2020). Regulating human control over autonomous systems. *Regulation & Governance*, *15*(4), 1071–1091. https://doi.org/10.1111/rego.12344

*Fleming, P. (2019). Robots and organization studies: Why robots might not want to steal your job. *Organization Studies*, *40*, 23–38.

*Flyverbom, M., Deibert, R., & Matten, D. (2019). The governance of digital technology, big data, and the internet: New roles and responsibilities for business. *Business & Society*, *58*, 3–19.

*Gellert, R. (2020). Comparing definitions of data and information in data protection law and machine learning: A useful way forward to meaningfully regulate algorithms? *Regulation & Governance*, *16*(1), 156–176. https://doi.org/10.1111/rego.12349

Girish, D. (2020, September 9). "The Social Dilemma" review: Unplug and run. *The New York Times*. https://www.nytimes.com/2020/09/09/movies/the-social-dilemma-review.html

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, *14*(2), 627-660.

*Gregory, R. W., Henfridsson, O., Kaganer, E., & Kyriakou, H. (2020). The role of artificial intelligence and data network effects for creating user value. *Academy of Management Review*, *46*, 534–551.

*Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California Management Review*, *61*(4), 5–14.

*Hartmann, P., & Henkel, J. (2020). The rise of corporate science in AI: Data as a strategic resource. *Academy of Management Discoveries*, *6*, 359–381.

Heukamp, F. (2020). AI and the leadership development of the future. In C.J. F. Heukamp (Eds.), *The future of management in an AI world: Redefining purpose and strategy in the fourth industrial revolution* (pp. 137–148). Palgrave Macmillan.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*(5), 359–366.

Hosanagar, K. (2020). *A human's guide to machine intelligence: How algorithms are shaping our lives and how we can stay in control*. Penguin Books.

*Huang, M. H., Rust, R., & Maksimovic, V. (2019). The feeling economy: Managing in the next generation of artificial intelligence (AI). *California Management Review*, *61*(4), 43–65.

*Jago, A. S. (2019). Algorithms and authenticity. *Academy of Management Discoveries*, *5*, 38–56.

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, *3*, 305–360.

*Johnson, C. D., Bauer, B. C., & Niederman, F. (2019). The automation of management and business science. *Academy of Management Perspectives*, *35*(2), 292–309.

Kane, G. C., Young, A. G., Majchrzak, A., & Ransbotham, S. (2021). Avoiding an oppressive future of machine learning: A design theory for emancipatory assistants. *MIS Quarterly*, *45*, 371–396.

*Kang, C., & Frenkel, S. (2018, April 4). Facebook says Cambridge Analytica harvested data of up to 87 million users. *The New York Times*. https://www.nytimes.com/2018/04/04/technology/mark-zuckerberg-testify-congress.html

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, *62*, 15–25. https://doi.org/10.1016/j.bushor.2018.08.004

*Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, *14*, 366–410.

Khanna, T. (2018). When technology gets ahead of society. *Harvard Business Review*, *96*(4), 86–95.

*Koo, W. W., & Eesley, C. E. (2021). Platform governance and the rural–urban divide: Sellers' responses to design change. *Strategic Management Journal*, *42*(5), 941–967.

*Kumar, V., Rajan, B., Venkatesan, R., & Lecinski, J. (2019). Understanding the role of artificial intelligence in personalized engagement marketing. *California Management Review*, *61*(4), 135–155.

*Lambrecht, A., & Tucker, C. (2019). Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads. *Management Science*, *65*(7), 2966–2981.

*Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, *5*(1), 2053951718756684.

*Leicht-Deobald, U., Busch, T., Schank, C., Weibel, A., Schafheitle, S., Wildhaber, I., & Kasper, G. (2019). The challenges of algorithm-based HR decision-making for personal integrity. *Journal of Business Ethics*, *160*, 377–392.

*Leonardi, P. M., & Treem, J. W. (2020). Behavioral visibility: A new paradigm for organization studies in the age of digitization, digitalization, and datafication. *Organization Studies*, *41*(12), 1601–1625. https://doi.org/10.1177/0170840620970728

Lessig, L. (1998). The new Chicago school. *The Journal of Legal Studies*, *27*(S2), 661–691.

Lessig, L. (2006). *Code: Version 2.0*. New York. Basic Books.

*Lévesque, M., Obschonka, M., & Nambisan, S. (2020). Pursuing impactful entrepreneurship research using artificial intelligence. *Entrepreneurship Theory and Practice*, 1042258720927369.

Levi-Faur, D. (2012). From “big government” to “big governance.” In D. Levi-Faur (Ed.), *The Oxford handbook of governance* (pp. 3–18). Oxford University Press.

*Li, X. B., & Sarkar, S. (2014). Digression and value concatenation to enable privacy-preserving regression. *MIS Quarterly*, *38*(3), 679–698.

Libecap, G. D. (1989). *Contracting for property rights*. Cambridge University Press.

*Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90–103.

*Lundvall, B. Å. (2017). Is there a technological fix for the current global stagnation? A response to Daniele Archibugi, Blade Runner economics: Will innovation lead the economic recovery? *Research Policy*, *46*, 544–549.

*Martin, K. (2019). Ethical implications and accountability of algorithms. *Journal of Business Ethics*, *160*(4), 835–850.

*Metcalf, L., Askay, D. A., & Rosenberg, L. B. (2019). Keeping humans in the loop: Pooling knowledge through artificial swarm intelligence to improve business decision making. *California Management Review*, *61*(4), 84–109.

*Miklós-Thal, J., & Tucker, C. (2019). Collusion by algorithm: Does better demand prediction facilitate coordination between sellers? Management Science, *65*(4), 1552–1561.

Mozur, P. (2019, April 14). One month, 500,000 face scans: How China is using A.I. to profile a minority. *The New York Times*. https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html

*Munoko, I., Brown-Liburd, H. L., & Vasarhelyi, M. (2020). The ethical implications of using artificial intelligence in auditing. *Journal of Business Ethics*, *167*, 209–234.

*Murray, A., Rhymer, J., & Sirmon, D. G. (2020). Humans and technology: Forms of conjoined agency in organizations. *Academy of Management Review*, *46*, 552–571.

*Newlands, G. (2021). Algorithmic surveillance in the gig economy: The organization of work through Lefebvrian conceived space. *Organization Studies*, *42*(5), 719–737.

Newman, A. (2012). The governance of privacy. In D. Levi-Faur (Ed.), *The Oxford handbook of governance* (pp. 599–610). Oxford University Press.

*Newman, D. T., Fast, N. J., & Harmon, D. J. (2020). When eliminating bias isn’t fair: Algorithmic reductionism and procedural justice in human resource decisions. *Organizational Behavior and Human Decision Processes*, *160*, 149–167.

North, D. C. (1990). A transaction cost theory of politics. *Journal of Theoretical Politics*, *2*(4), 355–367.

*Nunan, D., & Di Domenico, M. (2017). Big data: A normal accident waiting to happen? *Journal of Business Ethics*, *145*, 481–491.

*Overgoor, G., Chica, M., Rand, W., & Weishampel, A. (2019). Letting the computers take over: Using AI to solve marketing problems. *California Management Review*, *61*(4), 156–185.

*Pachidi, S., Berends, H., Faraj, S., & Huysman, M. (2021). Make way for the algorithms: Symbolic actions and change in a regime of knowing. Organization Science, *32*(1), 18–41.

*Parsons, C. (2019). The (in) effectiveness of voluntarily produced transparency reports. *Business & Society*, *58*, 103–131.

Phan, P., Wright, M., & Lee, S.-H. (2017). Of robots, artificial intelligence, and work. *Academy of Management Perspectives*, *31*(4), 253–255. https://doi.org/10.5465/amp.2017.0199

Platt, E. (2020, December 1). S&P Global's $44bn deal shows data is the oil of the 21st century. *Financial Times*. https://www.ft.com/content/cd99579c-e01f-4a71-a124-e9c03598e5b9

Potier, L. (2019, September 8). New tech to help disabled people. *The Guardian*. https://www.theguardian.com/technology/2019/sep/08/the-five-technology-to-help-disabled-people-blindness-paralysis-research-ai.

Purdy, M. (2020, October 27). *Unlocking AI's potential for social good*. https://hbr.org/2020/10/unlocking-ais-potential-for-social-good.

*Rahman, H. A. (2021). The invisible cage: Workers' reactivity to opaque algorithmic evaluations. *Administrative Science Quarterly*, *66*(4), 945–988.

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation-augmentation paradox. *Academy of Management Review*, *46*(1), 192–210.

Rao, A., & Verweij, G. (2017). *Sizing the prize: What's the real value of AI for your business and how can you capitalise*. PwC.

Ratcliffe, R. (2020, November 30). Facebook and YouTube accused of complicity in Vietnam repression. *The Guardian*. https://www.theguardian.com/world/2020/dec/01/facebook-youtube-google-accused-complicity-vietnam-repression.

*Raveendhran, R., & Fast, N. J. (2021). Humans judge, algorithms nudge: The psychology of behavior tracking acceptance. *Organizational Behavior and Human Decision Processes*, *164*, 11–26.

*Redden, J. (2018). Democratic governance in an age of datafication: Lessons from mapping government discourses and practices. *Big Data & Society*, *5*(2), 2053951718809145.

*Rietveld, J., Ploog, J. N., & Nieborg, D. B. (2020). The coevolution of platform dominance and governance strategies: Effects on complementor performance outcomes. *Academy of Management Discoveries*, *6*, 488–513.

Romm, T., Zakrzewski, C., & Lerman, R. (2020, October 7). House investigation faults Amazon, Apple, Facebook and Google for engaging in anti-competitive monopoly tactics. *The Washington Post*. https://www.washingtonpost.com/technology/2016/10/06/amazon-apple-facebook-google-congress/

*Schafheitle, S., Weibel, A., Ebert, I., Kasper, G., Schank, C., & Leicht-Deobald, U. (2020). No stone left unturned? Toward a framework for the impact of datafication technologies on organizational control. *Academy of Management Discoveries*, *6*, 455–487.

Scherer, A. G., & Voegtlin, C. (2020). Corporate governance for responsible innovation: Approaches to corporate governance and their implications for sustainable development. *Academy of Management Perspectives*, *34*, 182–208.

Scott, W. R. (2013). *Institutions and organizations: Ideas, interests, and identities*. SAGE.

Selznick, P. (1957). *Leadership in administration: A sociological interpretation*. Harper & Row.

Shepardson, D., & Somerville, H. (2019, March 5). *Uber not criminally liable in fatal 2018 Arizona self-driving crash: Prosecutors 2018*. Reuters. https://www.reuters.com/article/us-uber-crash-autonomous/uber-not-criminally-liable-in-fatal-2018-arizona-self-driving-crash-prosecutors-idUSKCN1QM2O8

*Shrestha, Y. R., Ben-Menahem, S. M., & Von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. *California Management Review*, *61*(4), 66–83.

Singer, N., & Isaac, M. (2020, January 29). Facebook to pay $550 million to settle facial recognition suit. *The New York Times*. https://www.nytimes.com/2020/01/29/technology/facebook-privacy-lawsuit-earnings.html

Stempel, J. (2020, June 2). *Google faces $5 billion lawsuit in U.S. for tracking "private" internet use*. Reuters. https://www.reuters.com/article/us-alphabet-google-privacy-lawsuit/google-faces-5-billion-lawsuit-in-u-s-for-tracking-private-internet-use-idUSKBN23933H

Sunstein, C. R. (2018). *Citizens. In #Republic divided democracy in the age of social media* (pp. 163–164). Princeton University Press.

*Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review*, *61*(4), 15–42.

*Tschang, F. T., & Almirall, E. (2020). Artificial intelligence as augmenting automation: Implications for employment. *Academy of Management Perspectives*, *35*, 642–659. https://doi.org/10.5465/amp.2019.0062

*Vergne, J. P. (2020). Decentralized vs. distributed organization: Blockchain, machine learning and the future of the digital platform. *Organization Theory*, *1*(4), 2631787720977052.

*West, S. M. (2019). Data capitalism: Redefining the logics of surveillance and privacy. *Business & Society*, *58*, 20–41.

*Whelan, G. (2019). Born political: A dispositive analysis of Google and copyright. *Business & Society*, *58*, 42–73.

*Yeung, K. (2018). Algorithmic regulation: A critical interrogation: Algorithmic regulation. *Regulation & Governance*, *12*(4), 505–523. https://doi.org/10.1111/rego.12158

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for the future at the new frontier of power*. Profile Books.

*Zwitter, A. (2014). Big data ethics. *Big Data & Society*, *1*(2), 205395171455925. https://doi.org/10.1177/2053951714559253

Author Biographies

Deepika Chhillar is a doctoral candidate in the Organizational Behavior and Strategy area at the Gies School of Business at the University of Illinois, Urbana–Champaign. She received her Masters in Economics and BE in Computer Science from BITS Pilani. Prior to joining the Ph.D. program Deepika worked in various roles at Credit Suisse and Schlumberger. Deepika's research interests include understanding and enhancing the cultural life of organizations, studying technological discontinuities, and the governing role of institutions. Her dissertation work employs computational grounded thoery and natural language processing.

Ruth V. Aguilera (Harvard University, PhD) is the Distinguished Darla and Frederick Brodsky Trustee Professor in Global Business at the D'Amore-McKim School of Business at Northeastern University and a Visiting Professor at ESADE Business School in her native Barcelona. Her research lies at the intersection of strategic organization and global strategy, specializing in international corporate governance and corporate social responsibility/sustainability. She is currently an Associate Editor at the *Academy of Management Review*.