Journal of Business Ethics
https://doi.org/10.1007/s10551-022-05054-9

ORIGINAL PAPER

From Greenwashing to Machinewashing: A Model and Future Directions Derived from Reasoning by Analogy

Peter Seele[1] · Mario D. Schultz[2]

Received: 1 October 2020 / Accepted: 25 January 2022

Abstract

This article proposes a conceptual mapping to outline salient properties and relations that allow for a knowledge transfer from the well-established greenwashing phenomenon to the more recent machinewashing. We account for relevant dissimilarities, indicating where conceptual boundaries may be drawn. Guided by a "reasoning by analogy" approach, the article addresses the structural analogy and machinewashing idiosyncrasies leading to a novel and theoretically informed model of machinewashing. Consequently, machinewashing is defined as a strategy that organizations adopt to engage in misleading behavior (communication and/or action) about ethical Artificial Intelligence (AI)/algorithmic systems. Machinewashing involves misleading information about ethical AI communicated or omitted via words, visuals, or the underlying algorithm of AI itself. Furthermore, and going beyond greenwashing, machinewashing may be used for symbolic actions such as (covert) lobbying and prevention of stricter regulation. By outlining diverse theoretical foundations of the established greenwashing domain and their relation to specific research questions, the article proposes a machinewashing model and a set of theory-related research questions on the macro, meso, and micro-level for future machinewashing research. We conclude by stressing limitations and by outlining practical implications for organizations and policymakers.

Keywords Machinewashing · Greenwashing · AI ethics · Corporate political activity · Decoupling · Agency theory

Introduction

Automation, digitization, and machine learning have irreversibly entered the scene. However, at the moment, it remains unpredictable what impact, what level of disruption, and which threats and benefits digital technology will contribute to business and society. Hopes and fears, utopian and dystopian visions are equally discussed. Particularly, ethical questions arising from machine intelligence are prevalent and trending. But it is not all (ethical) gold that shines: Kevin Roose (2019) reported for the New York Times, from the World Economic Forum in 2019, describing the "hidden automation agenda of the Davos Elite" where business leaders were publicly praising and discussing ethical and "human-centered AI," whereas in private talks with other managers, consultants and investors, shared that "they are racing to automate their own workforces to stay ahead of the competition, with little regard for the impact on workers" (Roose, 2019).

Corporations developing or working with machine intelligence respond to this unease among humans and their politicians trying to calm the worries with ethics programs and the hiring of ethics experts to "navigate the moral hazards presented by artificial intelligence without press scandals, employee protests, or legal trouble" (Knight, 2019). Business consulting company KPMG, for example, named "AI ethicist" among the "[t]op 5 AI hires companies need to succeed in 2019" (Fisher, 2019). Hitherto, the recent exit of Google's Ethical AI co-leader casts at least doubt on the seriousness of such ethical engagement. Timnit Gebru, Google's former researcher, "said she was fired by the company after criticizing its approach to minority hiring and the biases built into today's artificial intelligence

✉ Mario D. Schultz
mario.schultz@usi.ch

Peter Seele
peter.seele@usi.ch

[1] Ethics and Communication Law Center (ECLC), USI Università della Svizzera italiana, via G. Buffi 13 (R.413), 6900 Lugano, Switzerland

[2] Ethics and Communication Law Center (ECLC), USI Università della Svizzera italiana, Via Buffi 13 (R.414), 6900 Lugano, Switzerland

Published online: 03 February 2022

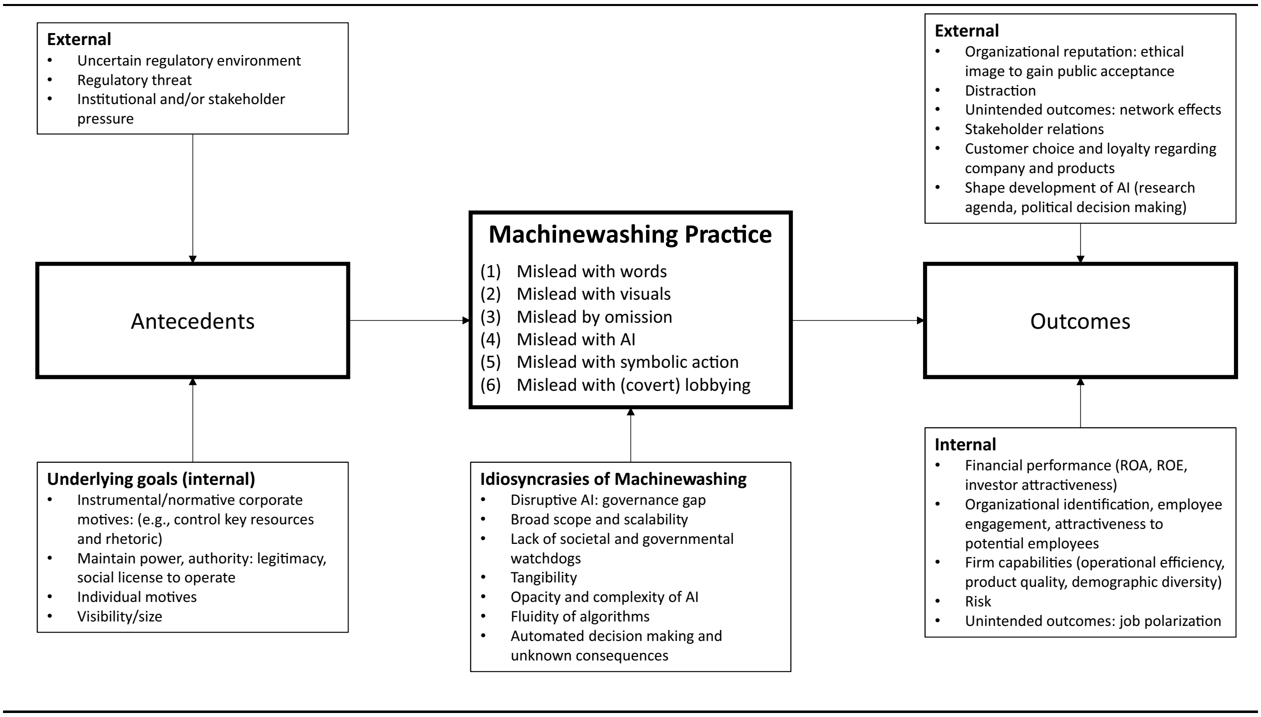
External
• Uncertain regulatory environment
• Regulatory threat
• Institutional and/or stakeholder pressure
Antecedents
Underlying goals (internal)
• Instrumental/normative corporate motives: (e.g., control key resources and rhetoric)
• Maintain power, authority: legitimacy, social license to operate
• Individual motives
• Visibility/size
Machinewashing Practice
(1) Mislead with words
(2) Mislead with visuals
(3) Mislead by omission
(4) Mislead with AI
(5) Mislead with symbolic action
(6) Mislead with (covert) lobbying
Idiosyncrasies of Machinewashing
• Disruptive AI: governance gap
• Broad scope and scalability
• Lack of societal and governmental watchdogs
• Tangibility
• Opacity and complexity of AI
• Fluidity of algorithms
• Automated decision making and unknown consequences
External
• Organizational reputation: ethical image to gain public acceptance
• Distraction
• Unintended outcomes: network effects
• Stakeholder relations
• Customer choice and loyalty regarding company and products
• Shape development of AI (research agenda, political decision making)
Outcomes
Internal
• Financial performance (ROA, ROE, investor attractiveness)
• Organizational identification, employee engagement, attractiveness to potential employees
• Firm capabilities (operational efficiency, product quality, demographic diversity)
• Risk
• Unintended outcomes: job polarization

Fig. 1 A model of Machinewashing

systems” (Metz & Wakabayashi, 2020). The media scandal that followed her exit and the firing of her co-head Margaret Mitchell has provoked broad skepticism about the actual reality behind corporate AI ethics programs (Johnson, 2021; Vincent, 2021). As a consequence, critical observers have drawn associations with a well-established business ethics concept, labeling AI ethics as “machinewashing”—derived from greenwashing as misleading environmental communication by companies (Obradovich et al., 2019). Now, what is machinewashing? The term has first been coined by Wagner (2018) and further refined by researchers from MIT Media Lab in a Boston Globe article, stressing that it represents a new form of greenwashing used to “[a]ddressing widespread concerns about the pernicious downsides of artificial intelligence (AI)—robots taking jobs, fatal autonomous-vehicle crashes, racial bias in criminal sentencing, the ugly polarization of the 2018 election—tech giants are working hard to assure us of their good intentions surrounding AI. But some of their public relations campaigns are creating the surface illusion of positive change without the verifiable reality. Call it “‘machinewashing’” (Obradovich et al., 2019).

Machinewashing is closely related to or used interchangeably with competing concepts that emerged in parallel: AI washing, (AI) ethics washing, ethical white washing, ethics bluewashing, and ethics theater (see Table 2). Given the evolving status of the discussion, research into the ethical issues raised by machinewashing is still in an early stage. Against this background, we reason by analogy to bring conceptual clarity into the machinewashing domain, bridging it with the established greenwashing domain (Cornelissen & Durand, 2014; Ketokivi et al., 2017; Vaughan, 2014). Thus, the guiding research questions of this article are: What are the antecedents, outcomes, idiosyncrasies, and underlying practices of machinewashing, where the source domain of greenwashing can inform the target domain of machinewashing (Ketokivi et al., 2017; Nersessian, 2008)? Furthermore, we ask: What are the similarities and differences, where the analogy does not fit, and where machinewashing goes beyond greenwashing given the disruptive momentum of AI? In other words, we analyze salient properties and relations that allow for transferring knowledge from the structural analogous greenwashing field. In addition, we stress the limitations of such a transfer in light of diverging theory assumptions that reveal unique characteristics of machinewashing and its distinct conceptual boundaries (Alvesson & Sandberg, 2011). Consequently, we summarize a conceptual model of machinewashing (see Fig. 1; Table 3) as a guiding framework that helps to organize and describe the various understandings and implications of machinewashing.

Although a scientific concept definition of machinewashing may help better understand its underlying practices and trigger focused (empirical) research, few attempts have been made to provide a concept definition (Nersessian, 2008). Therefore, *machinewashing is defined as a strategy that organizations adopt to engage in misleading behavior (communication and/or action) about ethical*

Artificial Intelligence (AI)/algorithmic systems. Machinewashing involves misleading information about ethical AI communicated or omitted via *words, visuals, or the underlying algorithm of AI itself. Furthermore, and going beyond greenwashing, machinewashing may be used for symbolic actions such as (covert) lobbying and prevention of stricter regulation.*

Given the relevance and strength of the link between the greenwashing and machinewashing domain, we outline a research program that builds on the analogy-derived model as a foundation (Ketokivi et al., 2017). As greenwashing occurred in a range of literatures, we show how the analogy links to multiple domains and opens the way for empirical study of the diverse machinewashing practices (Table 4 and 5). Thus, we provide evidence of the analogy's structural soundness and factual validity, encouraging future research and interdisciplinary dialog about machinewashing with a set of possible research questions on the macro, meso, and micro-level (Ketokivi et al., 2017). We emphasize that the suggested theory domains should not be seen as an exhaustive list but as an initial starting point to advance our understandings of machinewashing and broaden the scope of the research program (Alvesson & Sandberg, 2011; Ketokivi et al., 2017). Finally, we discuss broader avenues for future research and practical implications.

Conceptual Comparison of Greenwashing and Machinewashing

Analogies are central to management and organizational theories (Cornelissen & Durand, 2014; Ketokivi et al., 2017; Swedberg 2014; Vaughan, 2014). Many important theories that evolved in previous decades build on analogies to introduce new ideas, provide explanations and comprehensive insights into complex subjects, and trigger scientific discourse (Astley & Zammuto, 1992; Cornelissen, 2005; Williamson, 1971). An analogy typically links a source and a target domain through structural mapping, thereby utilizing the knowledge of the source to draw implications in the target domain (Ketokivi et al., 2017). In the following, we reason by analogy to connect the machinewashing domain with the source domain of greenwashing. To do so, we invoke a structural mapping that connects and transfers elements from the greenwashing debate to corresponding elements of the machinewashing discourse to ultimately arrive at a new and theoretically informed model of machinewashing (Gentner & Smith, 2012; Ketokivi et al., 2017; Nersessian, 2008; Vaughan, 2014). It is essential to be aware of the potential limits of such a knowledge transfer, given the distinct characteristics of the machinewashing phenomenon. A helpful way to approach this challenge is Alvesson and Sandberg's (2011) problematization strategy to identify and contest core assumptions underlying a given domain.[1] Thus, we use the problematization approach as an important means to stress the boundaries of the analogy, bringing forth salient elements of the machinewashing phenomenon.

Greenwashing and Machinewashing: State-of-the-Art and Core Assumptions

Greenwashing

In 1986, the biologist Jay Westerveld coined the term greenwashing, referring to the hotel industry's misleading practice of promoting the reuse of towels to save water and, thus, to conserve planetary resources (Becker-Olsen & Potucek, 2013). Whereas corporate communication emphasized an environmentally friendly image, the actual business intention to reduce laundry costs and increase profits remained in the dark (Orange, 2010). In this regard, greenwashing emerged as a concept that discusses corporate activities and practices that make "an organization look more environmentally friendly than it actually is" (Becker-Olsen & Potucek, 2013, p. 1318). Since its inception, the greenwashing concept has gained substantial traction in business ethics and related literatures, with several authors identifying and refining its boundaries. Table 1 summarizes the various academic and non-academic definitions of greenwashing and shows that today's understandings of greenwashing have become broader in scope,[2] going beyond a single focus on "the unjustified appropriation of environmental virtue by a company to create a pro-environmental image" (Marciniak, 2010, p. 49), covering also societal aspects, "i.e. merely businesses claiming environmental credentials and other social contributions while continuing to generate excessive harms such as social costs, i.e. 'business as usual'" (Sheehy, 2014, p. 626). The underlying core assumptions of the greenwashing domain can be summarized as follows. Greenwashing is: (1) "focused on information disclosure decisions," (2) "assumed to be a deliberate strategy," (3) "conceived primarily as a corporate phenomenon," and (4) "usually assumed to be beneficial for firms and detrimental to society (environment)" (5) relates to social and environmental

[1] Within the limited space of this paper, we focus on the crucial second step of the problematization approach: the identification and articulation of core assumptions. For a full outline of all stages of the problematization strategy, chapter 5 of Alvesson and Sandberg (2013) may be considered.

[2] For the purpose of this research, we adopt a wide understanding of greenwashing, which goes beyond mere symbolic communication about environmental aspects, and also covers various social dimensions, as the more recent term "CSR washing" underlines (Pope and Wæraas 2016).

Table 1 Greenwashing concepts and definitions

Author (year)	Title	Concept/definition
Oxford English Dictionary (2012)	Greenwashing, n	"Disinformation disseminated by an organization so as to present an environmentally responsible public image; a public image of environmental responsibility promulgated by or for an organization, etc., but perceived as being unfounded or intentionally misleading"
Merriam-Webster Dictionary (2020)	Greenwashing	"practice of promoting environmentally friendly programs to deflect attention from an organization's environmentally unfriendly or less savoury activities"
Laufer (2003)	Social accountability and corporate Greenwashing	"[F]orms of disinformation from organizations seeking to repair public reputations and further shape public images"
Walker and Wan (2012, p. 227)	The harm of symbolic actions and green-washing: corporate actions and communications on environmental performance and their financial implications	"[A] strategy that companies adopt to engage in symbolic communications of environmental issues without substantially addressing them in actions […]"
Seele and Gatti (2017, p. 239)	Greenwashing revisited: in search of at typology and accusation-based definition incorporating legitimacy strategies	"[G]reenwashing as co-creation of an external accusation toward an organization with regard to presenting a misleading green message."
Bowen (2014, p. 33)	After greenwashing: symbolic corporate environmentalism and society	"Greenwashing is a special case of 'merely symbolic' in which firms deliberately manipulate their communications and symbolic practices so as to build a ceremonial façade"
Marciniak (2010, p. 49)	Greenwashing as an example of ecological marketing misleading practices	"[T]he unjustified appropriation of environmental virtue by a company to create a pro-environmental image."
Matejek and Gössling (2014, p. 572)	Beyond legitimacy: a case study in BP's "Green Lashing"	"[S]ymbolic actions may even eclipse substantive activities entirely, a phenomenon generally referred to as greenwashing, or window dressing"
Marquis et al. (2016, p. 483)	Scrutiny, norms, and selective disclosure: a global study of greenwashing	"[A] symbolic strategy whereby firms seek to gain or maintain legitimacy by disproportionately revealing beneficial or relatively benign performance indicators to obscure their less impressive overall performance"
Guo et al. (2017, p. 524)	A path analysis of greenwashing in a trust crisis among Chinese energy companies: the role of brand legitimacy and brand loyalty	"Greenwashing here refers to the integration of two corporate behaviors: poor environmental performance and positive communication about environmental performance"
Sheehy (2014, p. 626)	Defining CSR: problems and solutions	"[…] i.e. merely businesses claiming environmental credentials and other social contributions while continuing to generate excessive harms such as social costs, i.e. 'business as usual'"

issues (Bowen, 2014, p. 26). From the established greenwashing source domain, we will now turn toward the target domain of machinewashing, depicting similarities its unique characteristics.

Machinewashing

Has only recently entered the business ethics research agenda in light of the widespread implementation of AI systems and rising concerns over the adverse impacts of AI (Benkler, 2019). Although machinewashing and similar concepts (see Table 2) are increasingly gaining momentum, the academic discourse around them remains dispersed. Current discussions are located at the intercept of multiple fields, including AI/machine ethics, information ethics, digital ethics, bioethics, robot ethics, and international law and governance literature. Considering existing machinewashing definitions (Table 2) along with the above-listed assumptions of greenwashing, one can see strong similarities, but also differences, where alternative assumptions may emerge (Alvesson & Sandberg, 2011): Machinewashing is a practice that: (1) focuses on information disclosure in the form of communicative activities using ethical language directed at various stakeholders (e.g., disclosure of ethics principles and guidelines); (2) focuses on misleading actions (e.g., symbolic and lobbying activities); (3) is seen as a deliberate practice; (3) is understood as an organizational phenomenon closely related technology firms, but not exclusively; (4) assumed to be benefit organizations while being detrimental for society; (6) relates to AI issues. Given that AI or algorithmic system issues mark the crucial difference, where machinewashing and greenwashing assumptions depart, a few words are in order to detail these salient elements (Alvesson & Sandberg, 2011).

Idiosyncrasies of Machinewashing

Machinewashing should not be considered a mere extension of greenwashing. At its core, machinewashing relates to broader AI issues, which warrant treating machinewashing as a distinct phenomenon rather than a digital version of greenwashing (see Table 2). Thus, the presented conceptual analogy is informed by the following peculiarities that characterize machinewashing concerning these broader AI issues (Ketokivi et al., 2017; Nersessian, 2008).

Disruptive AI

AI topics touch the general public and business's surface only within a few years. With the substantial advancements of computer technologies and increased processing power of modern microchips, machine learning and other advanced algorithmic systems have been developed and deployed on a large scale throughout society (Appenzeller, 2017; Crawford & Calo, 2016). Thus, machinewashing rapidly emerged just as the digital transformation did and does, disrupting societal values and legal systems rather than gradually changing them, as in the case of greenwashing (Becker-Olsen & Potucek, 2013).

Broad Scope and Scalability

AI use cases are only limited by designers' imagination, as the wide deployment and scalability across many areas of daily life show. Neurosurgical procedures, facial recognition, and speech synthesis systems to identify and imitate humans, self-driving cars, chatbots that respond to consumer emotions, or pre-trial risk assessments in courtrooms are all examples of systems featuring AI (Chesney & Citron, 2019; Hao, 2019; Hao & Stray, 2019; Huang & Rust, 2021a; Yuste et al., 2017). Thus, given the broad range of use cases, the spectrum for potential machinewashing becomes much more expansive than in the case of greenwashing.

Lack of Societal and Governmental Watchdogs

Characteristic for the emergence of greenwashing were newly formed civil society organizations and NGOs that aimed to point out corporations' adverse social and environmental impacts and hold them accountable (Becker-Olsen & Potucek, 2013; Sheehy, 2014). Further, dedicated governmental "watchdog" institutions evolved in several jurisdictions (see, e.g., Department for Environment Food and Rural Affairs (DEFRA) 2010; Federal Trade Commission (FTC) 2012). In contrast, when it comes to AI issues and machinewashing, apart from a few examples (see, e.g., AlgorithmWatch and Opendatawatch) civil society organizations and dedicated machinewashing watchdogs are still lacking (Koene et al., 2019).

Tangibility of AI Issues

The greenwashing phenomenon emerged in light of increasing attention for highly visible social and environmental problems, such as sweatshop working conditions and environmental degradation (Beder, 2002; Lim & Phillips, 2008). These social and green issues have a 'feel' and visibility. In contrast, issues related to machinewashing are much more abstract. Challenges associated with AI, such as privacy, algorithmic biases, discrimination, job loss, power- and information asymmetries, are often not apparent at first glance (Rust & Huang, 2021; Seele et al., 2021; Theodorou & Dignum, 2020). In other words, an oil spill in the Gulf of Mexico draws much greater attention than a data privacy

Table 2 Machinewashing concepts and definitions

Author (year)	Title	Concept/definition
Wagner (2018, p. 1)	Ethics as an escape from regulation: from ethics-washing to ethics-shopping?	"[E]thics is presented as a concrete policy option. Striving for ethics and ethical decision-making it is argued, will make technologies better. […] Unable or unwilling to properly provide regulatory solutions, ethics is seen as the 'easy' or 'soft' option which can help structure and give meaning to existing self-regulatory initiatives. In this world, 'ethics' is the new 'industry self-regulation'"
Obradovich et al. (2019)	Beware corporate 'machinewashing' of AI	Today, we may be witnessing a new kind of greenwashing in the technology sector. Addressing widespread concerns about the pernicious downsides of artificial intelligence (AI)—robots taking jobs, fatal autonomous-vehicle crashes, racial bias in criminal sentencing, the ugly polarization of the 2018 election—tech giants are working hard to assure us of their good intentions surrounding AI. But some of their public relations campaigns are creating the surface illusion of positive change without the verifiable reality. Call it "machinewashing"
Bietti (2020, p. 210)	From ethics washing to ethics bashing: a view on tech ethics from within moral philosophy	"[T]he term has been used by companies as an acceptable façade that justifies deregulation, self-regulation or market driven governance, and is increasingly identified with technology companies' self-interested adoption of appearances of ethical behavior"
McMillan and Brown (2019, p. 1)	Against ethical AI	In reference to Wagner (2018) it is summarized: "ethics washing is the use of working groups, guidelines, and manifestos as a counterbalance to calls for legal and regulatory frameworks which would ensure the safety of the public."
Rességuier and Rodrigues (2020, p. 2)	AI ethics should not remain toothless! A call to bring back the teeth of ethics	"Using ethics to prevent the implementation of legal regulation that is actually necessary is a serious and worrying abuse and misuse of ethics"
Floridi (2019, p. 186)	Translating principles into practices of digital ethics: five risks of being unethical	"Digital ethics shopping = def. the malpractice of choosing, adapting, or revising ('mixing and matching') ethical principles, guidelines, codes, frameworks, or other similar standards (especially but not only in the ethics of AI), from a variety of available offers, in order to retrofit some pre-existing behaviours (choices, processes, strategies, etc.), and hence justify them a posteriori, instead of implementing or improving new behaviours by benchmarking them against public, ethical standards"
Floridi (2019, p. 187)	Translating principles into practices of digital ethics: five risks of being unethical	"Ethics bluewashing = def. the malpractice of making unsubstantiated or misleading claims about, or implementing superficial measures in favour of, the ethical values and benefits of digital processes, products, services, or other solutions in order to appear more digitally ethical than one is."
Floridi (2019, p. 188)	Translating principles into practices of digital ethics: five risks of being unethical	"Digital ethics lobbying = def. the malpractice of exploiting digital ethics to delay, revise, replace, or avoid good and necessary legislation (or its enforcement) about the design, development, and deployment of digital processes, products, services, or other solutions"
Floridi (2019, p. 191)	Translating principles into practices of digital ethics: five risks of being unethical	"Ethics shirking = def. the malpractice of doing increasingly less 'ethical work' (such as fulfilling duties, respecting rights, and honouring commitments) in a given context the lower the return of such ethical work in that context is mistakenly perceived to be"

Table 2 (continued)

Author (year)	Title	Concept/definition
Coeckelbergh (2020, p. 4)	Green leviathan or the poetics of political liberty: navigating freedom in the age of climate change and artificial intelligence	"What if companies' insistence that they will develop AI for 'the earth' and use AI in a sustainable and climate-friendly way is just 'ethics washing', a fig leaf for doing business as usual?"
Yeung et al. (2020, p. 7)	AI governance by human rights-centred design, deliberation and oversight: an end to ethics washing	"It is hardly surprising that critics have dismissed these voluntary codes of conduct as 'ethics washing' given overwhelming evidence that the tech industry cannot be relied upon to honour its voluntary commitments"
Umbrello and van de Poel (2020, p. 21)	Mapping value sensitive design onto AI for social good principles	"[…] there is a danger that contribution that societal challenge and SDGs are used to for legitimisation of AI technologies that do not respect some fundamental ethical principles, i.e. there is a danger of ethical white-washing (which is already visible ta the webpages of some large companies)"
Metzinger (2019)	EU guidelines: ethics washing made in Europe	"Industry organizes and cultivates ethical debates to buy time – to distract the public and to prevent or at least delay effective regulation and policy-making."
Hao (2019)	In 2020, let's stop AI ethics-washing and actually do something	"We're falling into a trap of ethics-washing, where genuine action gets replaced by superficial promises"
Johnson (2019)	How AI companies can avoid ethics washing	"[E]thics washing—also called 'ethics theater'—is the practice of fabricating or exaggerating a company's interest in equitable AI systems that work for everyone. A textbook example for tech giants is when a company promotes 'AI for good' initiatives with one hand while selling surveillance capitalism tech to governments and corporate customers with the other"
Susser (2019)	Ethics alone can't fix big tech	"The result is "ethics theater"—or worse,"ethics washing"—a veneer of concern for the greater good, engineered to pacify critics and divert public attention away from what's really going on inside the A.I. sausage factories."
Kinstler (2020)	Ethicists aim to save tech's soul. Will anyone let them?	"[T]he practice of merely kowtowing in the direction of moral values in order to stave off government regulation and media criticism."
Waddell (2019)	The dangers of "AI washing"	"This "AI washing" threatens to overinflate expectations for the technology, undermining public trust and potentially setting up the booming field for a backlash."

leak triggered by a dysfunctional AI system. Consequently, also machinewashing becomes less palpable.

Opacity and Complexity of AI: Difficult to Grasp for Stakeholders

The abstract nature of issues related to the machinewashing phenomenon is partly due to the underlying complexity of AI systems (Hagendorff, 2019). Although some corporations make efforts to provide insights into their algorithms, their precise functioning is undoubtedly not evident, even for experts (Crawford & Calo, 2016; Jordan & Mitchell, 2015). AI systems remain black boxes for most stakeholders, often shielded by property rights (Martin, 2019; Noto La Diega, 2018). Consequently, the exact functionality of AI remains largely unknown and difficult to grasp, which makes opacity and complexity also crucial aspects of machinewashing.

Fluid Algorithms

AI-powered products or services are based on different forms of algorithms. As such, AI is difficult to grasp due to the fluid nature of the underlying algorithmic code (Buhmann et al., 2020). Especially machine learning system adapt their behavior to a context in real time (Yeung et al., 2020). Thus, rather than being associated with concrete or even tangible products as in greenwashing, AI can easily change shape via software updates, patches, or an altered deployment environment (Yeung, 2019). As a consequence, the machinewashing of AI products and services becomes more challenging to capture and follow.

Automated Decision Making and Unknown Consequences

At the core of AI systems is the possibility to automate processes and decision making with algorithms. Although AI, by definition, emerges from human design, its developing or deploying organization often slides into the background (Bryson, 2020). Further, algorithmic decision-making outcomes are not always what designers and organizations want or have planned for. Especially machine learning systems designed to develop independently and in real time carry the risk of triggering unintended or undesirable consequences (Wagner & Winkler, 2019). Thus, AI goes along with unpredictability and raises new responsibility questions when making independent or semi-automated (Coeckelbergh, 2020). In turn, the issue of unintended outcomes and the problem of responsibility translates to the machinewashing phenomenon.

In light of the depicted idiosyncrasies, the following paragraphs will structurally map the source and target domain to indicate fundamental relations that allow for a knowledge transfer between green and machinewashing (Gentner & Smith, 2012; Ketokivi et al., 2017; Nersessian, 2008; Vaughan, 2014). By building on five major dimensions, we stress the analogy and indicate where new ways of thinking are required to account for machinewashing idiosyncrasies (see Fig. 1; Table 3): (1) Antecedents (2) Underlying Goals (3) Practice (4) Examples and Manifestations (5) Outcomes.

Antecedents

Greenwashing

Greenwashing and machinewashing are both subject to antecedents that stem from the external and internal corporate environment. As depicted by Delmas and Burbano (2011), the external environment can be characterized as the nonmarket environment that includes the regulatory and broader public (e.g., the media), as well as the market environment (consumers, financiers, competitors). The internal environment includes organizational and individual psychological drivers that are summarized in Sect. (2) underlying goals. Previous research has identified the uncertain regulatory environment as a key driver of greenwashing, which indirectly touches on all other greenwashing drivers (Delmas & Burbano, 2011). Further, external drivers include pressure upon the corporation by activist groups, NGOs, and the media (Marciniak, 2010). Also, a company's institutional or operational context can be a decisive aspect in terms of prevalent industry norms and competitor practices (Du, 2014; Jones, 2012). Additionally, consumer and investor demands play a crucial role as drivers in the market environment (Delmas & Burbano, 2011; Sheehy, 2014).

Machinewashing

The uncertain regulatory context is also the most crucial institutional driver of corporate machinewashing in the nonmarket external environment (Benkler, 2019; Jobin et al., 2019; Wagner, 2018). This regulatory void has invited numerous organizations to launch soft-law initiatives making extensive but unsubstantiated ethical claims about their AI systems without having to risk any legal charges (Bietti, 2020; Jobin et al., 2019). Similar to greenwashing, in the nonmarket external environment, activist groups, NGOs, and the media play a role as monitors of machinewashing (Kinstler, 2020; McMillan & Brown, 2019). In terms of the external market environment, industry norms, competition, and consumers and investors are crucial external market dimensions to consider in the machinewashing debate. As pressure from these groups may trigger corporations to engage in symbolic communication about their AI practices (Mittelstadt, 2019).

Table 3 Machinewashing—model overview

Idiosyncrasies of Machinewashing	Machinewashing emerged rapidly along new and disruptive AI systems, challenging societal values and legal systems Broad range of AI use cases opens wide spectrum for machinewashing Lack of dedicated civil society and governmental watchdogs AI issues (privacy, algorithmic biases, discrimination etc.) and machinewashing not tangible at first glance Opacity and complexity of AI difficult to grasp for stakeholders. Machinewashing can be hidden in AI black boxes Fluid algorithms can quickly change shape (software patches), making machinewashing difficult to capture Automated decision making and unknown consequences obscuring responsibility for unintended adverse outcomes
Antecedents	Nascent activism, NGO, and media attention Uncertain regulatory environment; regulatory pressure
Underlying goals	Instrumental/normative corporate motives Reputation, competitive advantage Legitimacy, social license to operate Individual motives Firm visibility/size Maintain power, authority Control key resources (algorithms, data) and rhetoric
Practice*	Misleading communication gesture accompanied with symbolic action and open/covered corporate political activity (on multiple levels: legislative, judicial, and academic lobbying)
Outcomes	External Ethical image Indicate connection and adherence to principles Prevention of regulation or justification for deregulation/self-regulation Unintended outcomes: network effects Distract from major issues related to core business Internal Appropriation of (abstract) ethical virtues Financial/image gain Firm capabilities (operational efficiency, product quality, demographic diversity) Risk Unintended outcomes: such as job polarization
Definition	*Machinewashing is defined as a strategy that organizations adopt to engage in misleading behavior (communication and/or action) about ethical Artificial Intelligence (AI) / algorithmic systems. Machinewashing involves misleading information about ethical AI communicated or omitted* via *words, visuals, or the underlying algorithm of AI itself. Furthermore, and going beyond greenwashing, machinewashing may be used for symbolic actions such as (covert) lobbying and prevention of stricter regulation*

*For a detailed overview of machinewashing types and practices, see Table 4

Underlying Goals

Greenwashing

Underlying goals may be instrumental or normative in nature and relate to the firm's response to the external environment and depend on the internal organizational environment, including firm characteristics, incentive structures, and the ethical climate (Delmas & Burbano, 2011). Benefits arising from a green image include reputational gains and a potential competitive advantage and increased willingness to pay, mainly connected to large and highly visible firms (Matejek & Gössling, 2014; Szabo & Webster, 2020). Moreover, business legitimacy to build or maintain what has been termed the social license to operate remains a crucial underlying goal (Suchman, 1995; Walker & Wan, 2012). Moreover, organizational and individual psychological characteristics and the bounded rationality of agents can help explain greenwashing goals (Delmas & Burbano, 2011; Verbeke & Greidanus, 2009; Walker & Wan, 2012).

Machinewashing

Machinewashing follows similar goals. The AI ethics debate stresses the corporate quest for reputational gains, a competitive advantage, and business legitimacy (Benkler, 2019; Bietti, 2020; Douek, 2019; Koene et al., 2019). Additionally, technology corporations are driven by the incentive to maintain power and authority, triggering increased engagement in machinewashing (Kalluri, 2020). Given that algorithms and data are crucial to corporate success, large-scale and highly visible technology corporations have a general interest in maintaining control over these resources and the broader AI ethics debate (Floridi, 2019; Zuboff, 2019). Analogous to greenwashing, individual agents, in the form of managerial directors, may follow bounded rational agendas, triggering machinewashing (Papazoglou, 2019; Roose, 2019; Satariano, 2020).

Practice

Greenwashing

Greenwashing represents a symbolic communication gesture to create an unfounded or misleading pro-environmental or pro-social image (Laufer, 2003; Marciniak, 2010). The primary actor is the corporation making false or misleading claims about environmental and social attributes of products, services, and the organization's social and environmental performance. Whether a claim or corporate action is deceptive also depends on the audience's overall impression, as "greenwash is truly in the eye of the beholder" (Lyon & Montgomery, 2015, p. 6). To further specify greenwashing practices, it is helpful to look into existing regulations. Several types of false or misleading claims are the focus of regulatory standards and practice guidelines, issued by the U.S. Federal Trade Commission (FTC) (2012), the Australian Competition and Consumer Commission (2011), the Canadian Standards Association (CSA) (2008), and the UK Department for Environment, Food and Rural Affairs (DEFRA) (2010). Next to these public institutions, several NGOs and public/private sector organizations provide indicators that help to identify greenwashing: the Greenpeace Greenwash Criteria, the Greenwashing Index of EnviroMedia and the University of Oregon, the Seven sins of greenwashing by Terrachoice, and Futerra's signs of greenwashing (Zanasi et al., 2017). The most detailed outline of greenwashing types is provided by Lyon and Montgomery (2015), listing 11 mechanisms, also covering greenwashing as a nonmarket strategy that is (astroturf) lobbying. Summarizing these sources, greenwashing practices can be distinguished according to different categories or types of misleading behavior (see Table 4):

(1) Misleading with words, which may include (a) misleading claims (b) inaccurate claims (c) vague/unprovable claims (d) meaningless claims (d) overstatements/exaggerations
(2) Misleading with visuals or graphics. This aspect relates to various forms of visual rhetoric and semiotics. It also covers false or misleading seals, certifications, and labels.
(3) Misleading by omission. Noteworthy in this regard are corporate practices that relate to (a) complete omission of information, (b) selective disclosure, (c) incomplete comparisons (d) masking of information.
(4) Mislead with symbolic action refers to an inconsistency between promises about, e.g., social and environmental initiatives and actual actions in this regard.
(5) Mislead with (covert) lobbying. This category involves open or covert nonmarket activities aimed at favorable laws and regulations.

Machinewashing

Machinewashing practices match the above types of misleading behavior (see Table 4) and are further informed by the machinewashing idiosyncrasies (see Fig. 1; Table 3). Machinewashing includes the use of vague, inaccurate, and meaningless claims as well as jargon and exaggerations. Corporations may thereby not only communicate or mislead with words but also visuals and graphics. This involves the deceptive use of commercials, show robots but also covers certifications invoking unjustified commitments. Besides, it is not only the communication that may mislead but also what is omitted. Several omission types can be distinguished (see Table 4): complete omission of information, selective disclosure, incomplete comparisons, and information masking. In addition, AI itself may be used to mislead when imitating humans, presenting biased information, or obscuring responsibilities (Chesney & Citron, 2019; Wagner & Winkler, 2019; Yeung, 2019). This may involve deceptive corporate practices, such as strategically using property rights protection or function creep, when silently changing the algorithmic use case with a software update (Noto La Diega, 2018; Yeung, 2019). The misleading communication gestures are often accompanied by symbolic and aspirational measures and corporate political actions (Roose, 2019; Yeung et al., 2020). Machinewashing can represent a profound nonmarket strategy involving the full-level lobbying spectrum directed at the government, civil society, and academia (Bietti, 2020; Floridi, 2019; Ochigame, 2019; Rességuier & Rodrigues, 2020). Consequently, machinewashing practices go further than greenwashing, concerning nonmarket actions, but share the crucial similarity that deceptive claims and actions are subject to the audience's interpretation, resting "in the eye of the beholder" (Floridi, 2019; Lyon & Montgomery, 2015, p. 6).

Examples and Manifestations

Greenwashing

Greenwashing examples and manifestations are diverse and may be located at the product or firm level as listed in Table 4. One may find misleading claims such as

Table 4 Types of Misleading Greenwashing and Machinewashing Behavior

Type	Description	Greenwashing Example	Machinewashing Example
(1) Mislead with words			
(a) Misleading/vague claims	Broad claims without any specific meaning	Eco-friendly, environmental-friendly, eco-safe, all-natural, non-toxic, eco-conscious (see, e.g., Futerra and Terrachoice greenwash criteria in Zanasi et al., 2017)	Ethical AI, explainable AI, fair AI, trustworthy AI, human-friendly AI, sustainable AI, AI to benefit everyone
(b) Inaccurate claims	Claims or data that are wrong or made-up (closely related to 3 (a) complete omission of information)	"Common examples are tissue products that claim various percentages of post-consumer recycled content without providing any evidence" (TerraChoice, 2010, p. 10)	"IBM Watson is helping doctors outthink cancer, one patient at a time." […]'IBM needs to be held accountable for the image that it's producing of its successes compared to what they're actually able to deliver, because at a certain point it becomes an ethical issue… You're telling cancer patients that they should have a higher feeling of hope about their outcome and then under-delivering on that—to me, that's just dirty."
(c) Jargon claims	Claims that use language, terms, or jargon which do not resonate with stakeholders (especially customers)	Language and information that only an expert may understand (Futerra Sustainability Communications, 2009, p. 5)	Misleading and lengthy data and privacy policies, terms of service, and informed consent using legal and technical jargon (Obar & Oeldorf-Hirsch, 2020)
(d) Meaningless /irrelevant claims	Stressing a trivial ethical/ green aspect, whereas remaining business practices go against ethical or environmental standards	"For example, if a company brags about its boutique green R&D projects but the majority of spending and investment reinforces old, unsustainable, polluting practices." (Greenpeace Greenwash Criteria, in Zanasi et al., 2017, p. 65)	"YouTube has been driving millions of viewers to climate misinformation videos every day, a shocking revelation that runs contrary to Google's important missions of fighting misinformation and promoting climate action" (Corbin, 2020)
(e) Overstatements /exaggerations	Claims that make the organization or its products look better than they are. Overstatements claims that go far beyond the possibilities of the product or organization's capabilities	"For example, if a company were to do a million dollar ad campaign about a clean up that cost less" (Greenpeace Greenwash Criteria, in Zanasi et al., 2017, p. 65)	"If a typical person can do a mental task with less than one second of thought, we can probably automate it using AI either now or in the near future" (Ng, 2016)
(2) Mislead with visuals or graphics	Relates to images and video footage used in advertising, and seals, certifications and labels invoking unjustified commitments	"Green images that indicate a (un-justified) green impact e.g., flowers blooming from exhaust pipes" and "A label that looks like a third party endorsement … except it's made up" (Futerra Sustainability Communications, 2009, p. 5)	"Sophia is not the first show robot to attain celebrity status. Yet accusations of hype and deception have proliferated about the misrepresentation of AI to public and policymakers alike" (Sharkey, 2018)
(3) Misleading by omission			
(a) Complete omission of information	Claims that are made without proof of evidence (scientific confirmation)	"It could be right, but where's the evidence?" (Futerra Sustainability Communications, 2009, p. 5)	"IBM Watson is the Donald Trump of the AI industry—outlandish claims that aren't backed by credible data […] There is no way to validate what we're getting from IBM is accurate unless we test the real patients in an experiment" (Brown, 2017)

Table 4 (continued)

Type	Description	Greenwashing Example	Machinewashing Example
(b) Selective disclosure	Presented information creates a positive impression, while relevant information is kept back	"For example, if an industry or company has been forced to change a product, clean up its pollution or protect an endangered species, then uses PR campaigns to make such action look proactive or voluntary." (Greenpeace Greenwash Criteria, in Zanasi et al., 2017, p. 65)	Chatbots imitating humans and the challenge of consumers to know whether they are interacting with AI or not: "As this development gains traction, service providers have to decide whether to disclose the chatbot identity and, if so, whether to provide additional information about it. From an ethical viewpoint, withholding identity information does not prove tenable, as intransparency regarding the non-human chatbot identity may be perceived as deceptive and could be exploited by service providers" (Mozafari et al., 2020, p. 2916)
(c) Incomplete comparison	Basis for comparison is not provided	"Acme is more effective" (Lyon & Montgomery, 2015, p. 227)	"Ed Harbour, vice president of Implementation at IBM Watson […]"I believe very strongly Watson is ahead of the competition and we've got to continue to push [to make Watson better]. No, I don't think it's something that anybody can just do." (Brown, 2017)
(d) Masking of information	Relevant consequences of product/or services are omitted	"The ad leaves out or masks important information, making the green claim sound better than it is" (Greenwashingindex, in Zanasi et al., 2017, p. 66)	"A.I., most people in the tech industry would tell you, is the future of their industry, and it is improving fast thanks to something called machine learning. But tech executives rarely discuss the labor-intensive process that goes into its creation. A.I. is learning from humans. Lots and lots of humans. Before an A.I. system can learn, someone has to label the data supplied to it" (Metz, 2019)
(4) Mislead with AI*			
(a) (Ab)using AI to imitate humans	Refers to the use of AI to deceive/mislead consumers and, or the wider public	-	"Harmful lies are nothing new. But the ability to distort reality has taken an exponential leap forward with "deep fake" technology. This capability makes it possible to create audio and video of real people saying and doing things they never said or did" (Chesney & Citron, 2019, p. 1753)
(b) Biased real-time information	AI presents biased real-time information which cannot be verified by consumers	-	Discussing geographic information systems, Wagner and Winkler (2019, p. 7) note: "there is a considerable risk that users misinterpret the data provided to them and make bad decisions based on false or at best misleading information."

Table 4 (continued)

Type	Description	Greenwashing Example	Machinewashing Example
(c) Leaving AI code undisclosed	Using intellectual property rights law to avoid disclosure of algorithmic code (Related to mislead by omission)	-	Corporations may mislead about AI, leaving code undisclosed to avoid external assessment, referring to patent protection: “First, the overlap between, if not abuse of, intellectual property rights create a legal black box which is very difficult to open “ (Noto La Diega, 2018, p. 15)
(d) Hidden change of use case (function creep)	Using dynamic nature of algorithms (such as updates and patches) to adjust the mode of operation in the future	-	“Any change to the software of the system may affect the behaviour of the entire system or of individual components, extending their functionality, and these may change the system’s operational risk profile, including its capacity to operate in ways that might cause harm or violate human rights.” (Yeung, 2019, p. 63)
(e) Obscuring responsibility	Obscuring responsibility for unintended outcomes of semi-automated systems to human in the loop	-	“The collision of a Tesla car in semi-automated mode exemplifies the tendency to blame the proximate humans in the loop for unintended adverse consequences, rather than the surrounding socio-technical system in which the human is embedded. (Yeung, 2019, p. 61)
(5) Mislead with symbolic action			
(a) Policy practice gap	Refers to an inconsistency between promises about initiatives and actual actions	“Such as efficient light bulbs made in a factory which pollutes rivers” (Futerra Sustainability Communications, 2009, p. 5)	“The decisions of internal AI ethics committees are subjected to internal limits, subordinated to the endorsement of high management and dependent on company funding. This dependency on the company’s benevolence makes such efforts inadequate for addressing serious cases of company misconduct and also importantly unfit for achieving desirable policy outcomes” (Bietti, 2020, p. 216)
(b) Instrumentalization of ethics and moral philosophy*	Involves the instrumental use of ethics to achieve organizational outcomes	-	“[T]he trivialization of ethics and moral philosophy now understood as discrete tools or pre-formed social structures such as ethics boards, self-governance schemes or stakeholder groups” (Bietti, 2020, p. 210)
(6) Mislead with (covert) lobbying			
(a) Legislative lobbying	Involves open or covert non-market actions aimed at favorable laws and regulations	“For example, if advertising or public statements are used to emphasize corporate environmental responsibility in the midst of legislative pressure or legal action” (Greenpeace Greenwash Criteria, in Zanasi et al., 2017, p. 65)	“Google led the way with what would become one of the world’s richest lobbying machines. In 2018 nearly half the Senate received contributions from Facebook, Google and Amazon, and the companies continue to set spending records” (Zuboff, 2021)

Table 4 (continued)

Type	Description	Greenwashing Example	Machinewashing Example
(b) Academic lobbying*	Funding of research that favors corporate interests and helps to steer the academic debate	–	"Facebook has invested in the TU Munich – funding an institute to train AI ethicists. Similarly, until recently Google had engaged philosophers Joanna Bryson and Luciano Floridi for an 'Ethics Panel,' however this was abruptly discontinued at the end of last week. Had it not been for this, Google would have had direct access via Floridi, a member of HLEG AI, to the process by which this group will develop the political and investment recommendations for the European Union starting this month" (Metzinger, 2019)

*Applies specifically to machinewashing

eco-friendly, environmental-friendly, eco-safe, all-natural, non-toxic, eco-conscious in product descriptions (see, e.g., Futerra and Terrachoice greenwash criteria in Zanasi et al., 2017). Further, companies may use language and information that only an expert may understand or publish inaccurate, exaggerated, or meaningless claims (Greenpeace Greenwash Criteria, in Zanasi et al., 2017). Misleading visuals or graphics manifest as green-, eco-, or social labels that highlight that a product or service meets the respective standards of the label (Bowen, 2014). In this regard, greenwashing occurs when such certificates are erroneously adopted for products and services, which actually do not meet the certified standards.

Similarly, corporate communication may mislead by omission: presenting environmental product attributes as proactive achievements while missing to disclose that the action was a requirement to comply with existing regulations (Becker-Olsen & Potucek, 2013; Walter, 2010). Greenwashing can also manifest as symbolic action such as: "efficient light bulbs made in a factory which pollutes rivers" (Futerra Sustainability Communications, 2009, p. 5). Further, a green corporate image may be publicly presented, while the corporation lobbies against the environmental legislation (Walter, 2010).

Machinewashing

Among the most prominent expression of machinewashing (see Table 4) is the corporate adherence to ethical principles and values listed in guidelines or similar standards, which often communicate vague claims of explainable, human-friendly, sustainable, or trustworthy AI (Floridi, 2019; Jobin et al., 2019; Umbrello and van de Poel 2020; Yeung et al., 2020). Machinewashing also manifests as deceptive vague, and inaccurate claims, such as the example of IBM's flagship AI shows: "IBM Watson is helping doctors outthink cancer, one patient at a time" (Brown, 2017). Further, machinewashing may hide in the technical or legal jargon of lengthy data and privacy policies and service terms, which do not resonate with average consumers (Obar & Oeldorf-Hirsch, 2020). Corporations also mislead with visuals and graphics, as the show robot Sophia exemplifies: "Sophia is not the first show robot to attain celebrity status. Yet accusations of hype and deception have proliferated about the misrepresentation of AI to public and policymakers alike" (Sharkey, 2018). Similar to eco-labels, an increasing number of initiatives aim to introduce AI certification schemes, which run the same risk as the aforementioned, in case of the absence of adequate control (AI Ethics Impact Group, 2020; IEEE Standards Association, 2020). Misleading behavior may also involve AI itself, such as providing biased real-time information: "there is a considerable risk that users misinterpret the data provided to them and make bad decisions based on false or at best misleading information" (Wagner & Winkler, 2019, p. 7). Or by using misleading claims about patent protection to leave AI code undisclosed and avoid external assessment: "First, the overlap between, if not abuse of, intellectual property rights create a legal black box which is very difficult to open " (Noto La Diega, 2018, p. 15).

Further, symbolic actions involving corporate ethics boards, in-house philosophers, ethics working groups, and multi-stakeholder partnerships are prominent examples often associated with machinewashing. Particularly when these actions fail to produce meaningful results, as in the case of the instrumentalization of ethics to achieve organizational outcomes (Bietti, 2020; McMillan & Brown, 2019; Papazoglou, 2019). Another noteworthy expression of machinewashing is (covert) lobbying. Whereas legislative and judicial lobbying are well-known practices, lobbying academia to favor corporate interests is relatively new (Bietti, 2020; Rességuier & Rodrigues, 2020). Some have argued that the technology industry is manipulating academia to shape the AI debate by financing institutes that engage in AI ethics research (Ochigame, 2019; Ochigame et al., 2019). A recent New York Times article shows that this strategy focuses not only on soft-law research but also on hard-law and practices-oriented training directed at antitrust regulators and judges from multiple nations around the globe (Wakabayashi, 2020). As shown by Hagendorff and Meding (2020), such academic-industry cooperation is growing, with conflicts of interest remaining often undisclosed (see, e.g., Wright et al., 2018).

Outcomes

Greenwashing

Corporations engage in greenwashing practices to create a pro-social or pro-environmental image that may lead to reputational gains (Laufer, 2003; Marciniak, 2010). A favorable impression can also distract from an unsustainable core business (Marquis et al., 2016). In addition, several studies highlight the role of consumer behavior, particularly in light of perceived corporate greenwashing and the willingness to pay more for environmentally friendly products (Laroche et al., 2001; Nyilasy et al., 2014; Szabo & Webster, 2020). Another potential outcome of greenwashing is the potential financial gain and access to capital (Delmas & Burbano, 2011; Du, 2014). Past literature shows that firms are also exposed to risk in light of external stakeholders' backslash for greenwashing (Szabo & Webster, 2020). Further, research has shown that retaining human capital can be an important outcome (Ramdhony, 2018). Ambiguity and aspirational talk about CSR may benefit employee engagement, talent retention, and attraction, impacting firm capabilities and operational efficiency (Winkler et al., 2020).

Machinewashing

Machinewashing uses or misuses ethics to engineer or fabricate a public image or façade that boosts the organizational reputation to appease and gain the acceptance of the public and critical stakeholders (Floridi, 2019; Johnson, 2019). Machinewashing may also create unintended network effects when misleading information is automatically distributed at scale and real time (Wagner & Winkler, 2019). In this way, machinewashing may also impact customer choice of products and company loyalty but also help to distract and "retrofit some pre-existing behaviours (choices, processes, strategies, etc.)" (Floridi, 2019, p. 186). Further, machinewashing may shape AI's future development, given the underlying goal to prevent, avoid, counter, delay, revise or replace legislative efforts as a way to manage environmental risks (Floridi, 2019; Rességuier & Rodrigues, 2020). From an internal perspective, machinewashing may also relate to financial performance and impact the firm capabilities (McLennan et al., 2020; Rakova et al., 2020). A noteworthy example was Google dropping its AI ethics council after employee protests (Levin, 2019). The example also shows the potential impact of machinewashing on organizational identification, employee engagement, and an employer's attractiveness. Additionally, unintended outcomes may be triggered, such as increased job polarization, where promises of AI may manifest for high-skill workers but leave middle-skilled workers worse off (Dau-Schmidt, 2018; Huang & Rust, 2021b).

From the Structural Analogy Towards a Definition of Machinewashing

In the sections above, we reasoned by analogy to structurally map the machinewashing phenomenon and delineate its unique conceptual boundaries in a new model (Gentner & Smith, 2012; Ketokivi et al., 2017; Nersessian, 2008; Vaughan, 2014). Given the novelty of the phenomenon, few attempts have been made to provide a conceptual definition of machinewashing (see Table 2). Yet, a scientific concept definition may help to better understand the diverse machinewashing practices and trigger a focused research program (Nersessian, 2008). Thus, a definition can assist future attempts to operationalize and engage in the empirical study. This is crucial when it comes to identifying and measuring machinewashing and its outcomes. Further, from a practical perspective, a conceptual definition can assist corporations in avoiding misleading about their AI ethics activities. Consequently, we believe there is a need for a comprehensive theory-informed machinewashing definition. Therefore, the following definition derives from the structural analogy and the idiosyncrasies of the outlined machinewashing model (Ketokivi et al., 2017; Nersessian, 2008).

Definition: machinewashing is defined as a strategy that organizations adopt to engage in misleading behavior (communication and/or action) about ethical Artificial Intelligence (AI) / algorithmic systems. Machinewashing involves misleading information about ethical AI communicated or omitted via *words, visuals, or the underlying algorithm of AI itself. Furthermore, and going beyond greenwashing, machinewashing may be used for symbolic actions such as (covert) lobbying and prevention of stricter regulation.*

The Analogy as Foundation for Future Machinewashing Research

Machinewashing as a novel and multifaceted phenomenon requires the close attention of practitioners, regulators, and the broader public, making it a highly interesting and timely subject of study for business ethics and related fields (Bietti, 2020; Johnson, 2015; Martin et al., 2019). Given the relevance and strength of the link between the greenwashing and machinewashing domains established above, we argue that the analogy-derived machinewashing model can constitute the basis for future research to study the diverse machinewashing practices (Fig. 1; Table 4). Thus, in the following, we provide evidence of the structural soundness and factual validity of the analogy by demonstrating how machinewashing: (1) links to existing theories underpinning the greenwashing debate, and (2) lends itself to multiple research questions for empirical study (Ketokivi et al., 2017). The below-described theories are grouped according to macro, meso, and micro levels, indicating the respective focus of the theories (Cornelissen & Durand, 2014; Jeurissen, 1997). In light of Alvesson and Sandberg's (2013) problematization strategy, we emphasize the need to challenge the suggested theories concerning their underlying core assumptions and the distinct characteristics of machinewashing (see Idiosyncrasies of Machinewashing). Consequently, the proposed theories should not be seen as an exhaustive list but as an initial starting point to expand the analogy and broaden the scope of the machinewashing research (Table 5).

Macro-level Theories

Future machinewashing research may focus on macro theories dealing with organizations and the broader environments they are embedded in. This includes studying machinewashing concerning macro-level institutions, such as markets and governments, and more general cultural traditions of a given context (Cornelissen & Durand, 2014; Jeurissen, 1997). To adapt and shape this institutional environment, organizations may adopt machinewashing practices (Scott, 2014). In this regard, the following macro-level theories offer fruitful avenues to shed light on machinewashing practices: legitimacy theory, corporate political activity, and resource-dependence theory.

Legitimacy refers to the extent to which an organization operates according to a given set of social rules, such as formal and informal ones (Long & Driscoll, 2008; Suchman, 1995; Suddaby et al., 2017). *Legitimacy theory* follows the idea that organizations strive for social approval of their conduct in a given setting while avoiding disapproval to ensure their business continuity. Thus, to maintain what has been termed "social license to operate," organizations engage in a constant legitimization process (Melé & Armengou, 2016). Legitimacy theory can help shed light on machinewashing practices that aim at various types of legitimacy—pragmatic, moral, and cognitive (Bietti, 2020; Greenwood et al., 2011). For instance, previous research associates greenwashing with an organization's failure to attain pragmatic legitimacy (Seele & Gatti, 2017). Analog, this raises the question: how does machinewashing relate to different types of legitimacy—pragmatic, moral, and cognitive (Long & Driscoll, 2008)? Previous legitimacy research also associates corporate ethics codes with strategic self-interest, which may lead to symbolic isomorphism (Long & Driscoll, 2008). Similarly, today's AI ethics codes may serve the organizational self-interest and gradually converge, such that a hypocritical approach to ethics may become the new social norm in the institutional context (Jobin et al., 2019; Long & Driscoll, 2008). Thus, legitimacy theory can also illuminate mimetic pressure from the institutional environment regarding how mimetic pressure in the market might affect a firm's use of machinewashing. This also raises the question: which kind of pressure (mimetic or normative) is more influential in adopting machinewashing practices?

Future research may also benefit from the theoretical insights of *corporate political activity (CPA) research,* including *lobbying* and *public affairs.* As shown above, machinewashing is strongly associated with legislative, judicial, and academic lobbying (Bietti, 2020; Floridi, 2019; Ochigame, 2019; Rességuier & Rodrigues, 2020). Here, the CPA and public affairs lenses may help study machinewashing practices of organizations that are actively shaping their institutional environments. As discussed by den Hond et al. (2014), corporations may simultaneously play on two chessboards, the self-regulatory chessboard and the CPA chessboard, to achieve overall favorable outcomes in the non-market sphere. Technology organizations may similarly play on the AI ethics chessboard while engaging

Table 5 Macro, meso, and micro-level theories and questions to study machinewashing

Theory	Key assumptions	Examples of future research questions
Theories of organizations in their environments (Macro)		
Legitimacy theory	Organizations' long-term survival hinges on legitimacy: the conformity with societal norms (formal/informal)	How might mimetic pressures affect a firm's use of machinewashing? Which kind of pressure (mimetic or normative) is more influential in adopting machinewashing practices? How does machinewashing relate to different types of legitimacy—pragmatic, moral, and cognitive? How does machinewashing affect the credibility of an AI strategy—or the organization's reputation as such?
Corporate political activity/lobbying	Organizations as strategic players shaping the non-market environment	Does machinewashing distract society from questioning the limits of current AI ethics programs and from pushing governments to adopt stricter regulations? How are lobbying expenditures against more stringent AI regulations related to organizational spending on AI ethics programs? Which role do societal watchdogs or internal whistleblowers play in exposing a misalignment of the two chessboards? To what extent does organizational funding of public research favor corporate interests?? Does the lobbying of academic institutions undermine the independence of academia?
Resource-dependence theory	The long-term survival and growth of firms' hinges on access to critical resources from the external firm environment	How does resource pressure impact the adoption and use of machinewashing? How is AI used as a resource itself to engage in machinewashing and influence external stakeholders? Are AI ethics boards used to limit dependence on or gain access to critical resources?
Intermediate, organization-focused theories (Meso)		
Organizational Institutionalism	Organizations are embedded in institutional arrangements adapting to internal and external pressures	To what extent are AI ethics principles aligned with day-to-day organizational practices? How do internal procedures and the goal to preserve organizational efficiency relate to the adoption of machinewashing practices? What role can ethics boards play in ensuring that ethics guidelines and codes are translated to daily practice? To what extent are specific machinewashing practices (already) institutionalized in a given organization?

Table 5 (continued)

Theory	Key assumptions	Examples of future research questions
Instrumental and deliberative CSR	A discursive approach to organizations' responsibilities (e.g., discourse ethics, agnostic rhetoric, license to critique)	Can a deliberative approach to AI ethics offset the lack of societal watchdogs and assist in transferring principles into practice? Who influences the current AI ethics discourse, and why? In which way can organizational ethics boards contribute to the formation of ethics codes? How should an ethics board be structured to serve as an independent forum for discussion on weaknesses of AI? How are power and information asymmetries in the AI ethics discourse related to machinewashing practices? How can symbolic practices (e.g., ethics working groups and multi-stakeholder partnerships) be turned into credible and constructive discussions on ethical AI?
Signaling theory	Observing organizational behavior from an economic self-interest perspective or instrumental rationale	Does machinewashing pay off? Is machinewashing used to change perceptions about organizational AI ethics performance? Does evidence exist that the market values machinewashing? Does greater transparency about AI ethics programs, such as the disclosure of algorithmic code, mitigate information asymmetries between organizations and their stakeholders?
Theories of individuals within and around organizations (Micro)		
Agency theory	Issues arising when the principal employs an agent for value creation	(How) Does machinewashing impact agency cost? How can organizations control and verify that an agent acts in the principal's interest, not engaging in machinewashing practices? How do AI ethics programs relate to individual employees? What are the impacts of machinewashing on employee performance, well-being, and satisfaction? How can individual members of the organization be included in creating AI ethics codes and guidelines? How can individuals (continuously) challenge and be challenged by the code of their respective organizations? How can whistleblowers speaking out about weaknesses of corporate AI ethics be better protected/incentivized?

Table 5 (continued)

Theory	Key assumptions	Examples of future research questions
Attribution theory	Individuals 'attribution processes about organizational behavior	How do observers make sense of machinewashing communication? How is machinewashing related to purchasing and investment intentions and product as well as organizational loyalty? What micro-level attribution processes occur when consumers perceive AI ethics programs to be misleading? How do consumers and the wider public perceive deceptive practices, such as using intellectual property rights law to avoid disclosing algorithmic code or obscuring responsibility for unintended outcomes of semi-automated systems by blaming the human in the loop?

in the policymaking chess game to lobby against governmental regulations (Floridi, 2019; Yeung et al., 2020). However, organizations risk being exposed (machinewashing) as players on both chessboards, which may adversely impact their reputation and legitimacy if these two strategies are not aligned (den Hond et al., 2014; Rehbein et al., 2018). Interesting future research questions may revolve around the conceptual framing of two chessboards' logic, including institutionalization and paradoxes resulting from misalignment. Questions may include: Does machinewashing distract society from questioning the limits of current AI ethics programs and from pushing governments to adopt stricter regulations? How are lobbying expenditures against more stringent AI regulations related to organizational spending on AI ethics programs (Zuboff, 2021)? Which role do societal watchdogs or internal whistleblowers play in exposing a misalignment of the two chessboards? In addition, the CPA theory lens is beneficial in investigating the rising academic lobbying: to what extent does organizational funding of public research favor corporate interests? Does the lobbying of academic institutions undermine the independence of academia?

Resource-dependence theory (RDT) argues that firms' long-term survival and growth hinges on access to critical resources. Future research may build on this theoretical angle when studying machinewashing as a nonmarket organizational strategy used to "reduce environmental interdependence and uncertainty" (Hillman et al., 2009, p. 1404). Here, AI ethics and other related firm activities can be observed as mechanisms to reduce external dependence for critical resources and create a favorable non-market environment to secure essential resources from external stakeholders (Mellahi et al., 2016; Shirodkar et al., 2018)**.** From this perspective, the flow of digital (behavioral) data may be perceived as a crucial resource that machinewashing practices help to secure (Zuboff, 2019). Thus, interesting questions may relate to how resource pressure impacts the adoption and use of machinewashing? How is AI used as a resource itself to engage in machinewashing and influence external stakeholders? Previous research also highlights the value of RDT when understanding boards (Hillman et al., 2009). In particular, RDT may help to shed light on the way in which organizations may use AI ethics boards to limit dependence or gain certain resources, as the case of Google's ethics panel composition has shown (Metzinger, 2019).

Meso-level Theories

Organizational-focused theories can be utilized to better understand organizational antecedents and outcomes of machinewashing. Thus, organizational-focused theories are on an intermediate level between macro and micro-theories, opening the door to a better understanding of internal conditions under which organizations engage in machinewashing practices (Cornelissen & Durand, 2014). In this regard, future research may draw on insights from organizational institutionalism, instrumental/deliberative CSR, and signaling theory.

Organizational institutionalism. A fruitful avenue of future research is the decoupling concept as discussed in organizational institutionalism (Greenwood et al., 2018). Decoupling focuses on the potential gap between firms' policies and practices or, more precisely, means and ends (Bromley & Powell, 2012). Decoupling is relevant for machinewashing research, particularly to observe formal AI policies adopted by technology corporations concerning their daily business conduct. CSR and greenwashing research highlight the need to study talk action dynamics and, in light of aspirational CSR talk (Christensen et al., 2020a, 2020b; Glozer & Morsing, 2020)**.** Firms' resources dedicated to ethical AI may have limited or no relation to their intended AI goals. Thus, research questions along these lines include to what extent are organization AI ethics principles aligned with day-to-day practices? How do internal procedures and the goal to preserve organizational efficiency relate to the adoption of machinewashing practices? What role can ethics boards play in ensuring that ethics guidelines and codes are translated to daily practice? Another interesting question to explore is to what extent are specific machinewashing practices (already) institutionalized within a given organization?

Instrumental and deliberative CSR. Communication about ethical AI is often labeled as untrustworthy, aspirational, and not least as machinewashing (Benkler, 2019; B. Mittelstadt, 2019; Truby, 2020). Similar concerns have previously been raised about CSR as a mere marketing or public relations tool, where a substantial gap between symbolic and aspirational managerial 'talk' and the actual upholding of social responsibility standards prevails (Christensen et al., 2017; Haack et al., 2012; Winkler et al., 2020). In response to such critique, recent research observes how communication can move from an instrumental to a deliberative mode (B. D. Mittelstadt et al., 2015; Palazzo & Scherer, 2006; Scholz et al., 2019; Winkler et al., 2020). Seele and Lock (2015) outline a pathway about how idealistic principles can be translated into practice via discourse ideals. Further, Winkler et al. (2020) develop a framework of *agnostic rhetoric* that may facilitate the enactment of aspirational talk. In a similar vein, Christensen et al. (2017) emphasize a *license to critique*, outlining that ongoing communication (involving criticism and contestation) about sustainability standards is necessary to develop, adjust, and fine-tune standards as functional governance tools. Consequently, such deliberative approaches may also represent fruitful avenues for future machinewashing research, triggering important questions: Can a deliberative approach to AI ethics offset the lack of societal watchdogs and assist in transferring principles into practice? Who controls the current AI ethics discourse, and why? How are power and information asymmetries in the AI ethics discourse related to machinewashing practices? How can symbolic practices (e.g., ethics working groups and multi-stakeholder partnerships) be turned into meaningful and constructive discussions on ethical AI?

Signaling theory focuses on the behavior of a sender (e.g., a corporation) and a receiver (e.g., the customers, the public) in a communication process that is characterized by diverging information access (Connelly et al., 2011). Thus, signaling theory leans towards micro-levels, dealing with how the organization "signals" information to receivers, who choose the way of interpreting the communicated signal (Connelly et al., 2011). In greenwashing literature, the disclosure of corporate signals is studied as a form of deliberative communication of positive information that lacks observable attributes (Bowen, 2014). From this perspective, the corporate engagement in symbolic practices of 'ethical talk' about AI can be an effective signal toward society, underlining the corporate commitment to ethical values (Walker & Wan, 2012; Wu et al., 2020). In this regard, signaling theory stresses that corporations might signal false claims "if initial financial payoffs outweigh future losses once the truth becomes known" (Whelan & Demangeot, 2015, p. 1). Thus, signaling theory opens the door for observing machinewashing from an economic self-interest perspective and instrumental rationale. Analogous to greenwashing research, one may ask whether machinewashing pays off for an organization (Berrone et al., 2017)? Is machinewashing used to change perceptions about organizational AI ethics performance? Does evidence exist that the market values machinewashing (Du, 2014)? Does greater transparency about AI ethics programs, such as the disclosure of algorithmic code, mitigate information asymmetries between organizations and their stakeholders?

Micro-level Theories

Whereas the previous paragraphs outline machinewashing issues connected to organizations and the environments in which they are embedded, the study of machinewashing on a micro-level is likewise important (Lyon & Montgomery, 2015). Micro-level theories focus on the individual within and around organizations (Cornelissen & Durand, 2014). Micro-level theories are essential, as machinewashing practices can influence or be shaped by prominent individuals within an organization. In addition, as a phenomenon that lies in the beholder's eye, it is crucial to study how external stakeholders evaluate machinewashing. Thus, how individuals perceive various machinewashing practices.

Previous greenwashing research builds on two prominent theories in this regard: *agency* and *attribution theory*.

Agency theory may provide a fruitful avenue for future research to analyze internal antecedents of machinewashing**.** Individual actors may play a crucial role in designing, implanting, and executing AI ethics programs. Managers may engage in machinewashing activities and thereby follow personal interests (as opposed to shareholder interests), creating costs and inefficiencies for the principle (the organization) and broader society (Bosse & Phillips, 2016; Hadani & Schuler, 2013; Petrenko et al., 2016). An interesting question in this regard is how do managers instrumentalize AI ethics or build on information asymmetries to pursue their own career goals? Mark Zuckerberg's Facebook role has recently received critical attention (Abdalla & Abdalla, 2020; Zuboff, 2021). Thus, agency theory raises questions, such as: how does machinewashing impact agency cost? How can organizations control and verify that an agent acts in the principal's interest, not engaging in machinewashing practices? Further, it may be interesting to apply agency theory regarding employee's relation to machinewashing practices. The introductory example of Timnit Gebru and Margaret Mitchell may be an interesting starting point to study how machinewashing practices may impact organizational identification, employee engagement, and performance (Johnson, 2021; Vincent, 2021). Thus, further questions raised by agency theory include: how do AI ethics programs relate to individual employees? What is the impact of machinewashing on employee performance, well-being, and satisfaction? How can individual members of the organization be included in creating AI ethics codes and guidelines? How can individuals (continuously) challenge and be challenged by the code of their respective organizations?

Another theory that may prove helpful for future research is *attribution theory* (Harvey et al., 2014; Lange & Washburn, 2012). Attribution theory may help to capture external evaluations of machinewashing on an individual basis. This may include studying individuals' perceptions of machinewashing practices and the conclusions they draw from them. Thus, important avenues for future research may consist of how observers make sense of machinewashing communication and whether this may relate to aspects such as purchase and investment intention or product and organizational loyalty (Ginder et al., 2019; Pizzetti et al., 2019). In addition, attributional processes may help explain individuals' emotions and behavior when facing potential deceptive AI claims. This raises several questions: What micro-level attribution processes occur when consumers perceive AI ethics programs to be misleading? How do consumers and the wider public perceive deceptive practices, such as using intellectual property rights law to avoid disclosing algorithmic code or obscuring responsibility for unintended outcomes of semi-automated systems by blaming the human in the loop?

Empirical Inquiry

To further explore and theorize about machinewashing empirical inquiry is paramount, as current research remains largely conceptual (Benkler, 2019; Jobin et al., 2019; Wagner, 2018). As shown above, future research may utilize the theories from the three focus levels as a basis for empirical study. The suggested research questions (see Table 5) can be seen as initial ideas to apply these theories to different organizations, different organizational and institutional contexts, and investigate individual behavior. The complexity of machinewashing practices demands a holistic way of proceeding that draws on different methodological approaches, such as quantitative, qualitative, and combined methods. Further, to gain profound insights into machinewashing, it is necessary to collect data on multiple levels (individuals, organizations, and aggregated institutional contexts) and study processes and how they develop over time. The introductory case of Timnit Gebru exemplifies the involvement of multiple levels (from individual to aggregated institutions) and the need to study machinewashing practices from within and outside the organization and how it unfolds over time (Johnson, 2021; Vincent, 2021).

Future research may also benefit from further operationalizing the outlined machinewashing concept. This may include developing ways to measure stakeholder perceptions of machinewashing as a phenomenon that lies in the eye of the beholder (Lyon & Montgomery, 2015). In this regard, one possible pathway for future research is developing a scale that allows assessing perceptions of machinewashing. Scale development is a multistage process that helps create a validated scientific measurement instrument (DeVellis, 2017; Hinkin, 1995; Jian et al., 2014). In addition, scales can reveal insights into variables such as machinewashing that are not readily observable (DeVellis, 2017). Thus, a scale can benefit future research to generate and test hypotheses empirically, such as the perceived legitimacy of corporate AI soft policies.

Conclusion

With the widespread implementation of AI, organizations have started implementing AI ethics programs to meet societal and regulatory concerns and counter potential adverse outcomes of AI. However, critical voices increasingly question the reality behind corporate AI ethics, labeling it as machinewashing (Johnson, 2021; Ochigame, 2019). This article set out to offer conceptual clarity into the emerging machinewashing debate, discussing its resemblance and difference to greenwashing and providing a research program based on multiple theory lenses present in existing greenwashing literature. We have reasoned by analogy, invoking as structural mapping of greenwashing as the source and machinewashing as the target domain to allow for a knowledge transfer, thus, creating a theory-informed model of machinewashing (Ketokivi et al., 2017; Nersessian, 2008). Given the distinct idiosyncrasies of machinewashing, we have cautioned against a blind transfer of theoretical assumptions from greenwashing to machinewashing (Alvesson & Sandberg, 2013). Thus, we have emphasized the need for adaptations and critical reflection about greenwashing theories against distinct machinewashing characteristics. Further, we strongly encourage future research to draw on theoretical insights beyond the previously mentioned theories to allow for a rethinking and the generation of novel insights beyond the traditional "research box" or community (Alvesson & Sandberg, 2014). Overall, with our conceptual machinewashing model, the definition, and the theory-driven research program, we strive to trigger future research and cross-disciplinary dialog about machinewashing.

Limitations and Future Research

The introduction of the greenwashing-machinewashing analogy can enhance the understanding of organizational machinewashing practices and benefit theorizing in the emerging domain. However, researchers and theorists should not simply accept the analogy for its own sake. Therefore, an important future research direction is to critique the analogy, particularly in its boundary conditions (Ketokivi et al., 2017). In this regard, the section on machinewashing idiosyncrasies has shown where limitations of the analogy are situated, which in turn, can lead to novel insights. Thus, it is necessary to map and collect further evidence about the boundaries of the analogy to ensure the progress of the outlined research program. Another important starting point for future research to extend the themes highlighted in this manuscript relates to the broader discussion of whether AI systems or environmental issues go along with more systematic and universal impacts. For example, AI systems, with their possibility to automate work, decision making, idea production, value creation, organizational structures, and reward systems, increasingly disrupt the job market and firms' management processes (Chalmers et al., 2020; Lyytinen et al., 2017). This may trigger broader societal debates about aspects, such as the tension between increased automation and the possibility to uphold high employment rates. Future research may pay attention to such macro discussions in the CSR domain, where social and environmental issues are discussed concerning systemic perspectives on overconsumption and closed-loop economies (Kopnina, 2019; Murray et al., 2017).

Practical Implications

Machinewashing practices are difficult to assess, representing high credence issues similar to greenwashing. Over the past decades, new governance mechanisms and objective assessment methods emerged to approach challenges related to greenwashing and connected social and environmental problems (Beder, 2002; Lim & Phillips, 2008). With these approaches, regulators and NGOs are empowered to independently assess product and firm-level greenwashing by verifying product components or auditing work sites (Delmas & Burbano, 2011; see, e.g., Federal Trade Commission (FTC) 2012; TerraChoice, 2010). Regarding machinewashing, an objective external assessment is highly difficult or even impossible to perform without a firms' collaboration to share the algorithmic code or accept algorithmic auditing (Buhmann et al., 2020; Yeung, 2019). Thus, it remains a crucial challenge from a practical perspective to gain insights into and assess algorithmic black boxes (Martin, 2019). Theoretically, several solutions may be possible to tackle this challenge, ranging from voluntary transparency to complete disclosure requirements. Corporations may opt to transparently share the code of their algorithms with the public. However, in light of current market practices and corporate interests to protect business models and value chains, such an approach seems unlikely (Noto La Diega, 2018). A more feasible solution could be a state-level assessment body for algorithms, such as in the case of high-risk AI and the multistage certification process in the aviation industry (European Aviation Safety Agency, 2020) or an independent third-party verification performed by an auditing firm or delegated regulation unit as in co-regulation (Kamara, 2017). Above all, governments could implement new regulations for mandatory disclosure about algorithms and the handling of user data, similar to the transparency provided by soft- and hard-law approaches (Gatti et al., 2019) that aim at countering greenwashing.

Funding Open access funding provided by Università della Svizzera italiana.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Approval This article does not contain any studies with human participants performed by any of the authors.

References

Abdalla, M., & Abdalla, M. (2020). The Grey Hoodie Project: Big Tobacco, Big Tech, and the threat on academic integrity. *arXiv*. https://doi.org/10.1145/3461702.3462563

AI Ethics Impact Group. (2020). From principles to practice: n interdisciplinary framework to operationalise AI ethics. *VDE Association for Electrical Electronic & Information Technologies E.v., Bertelsmann Stiftung*. https://doi.org/10.11586/2020013

Alvesson, M., & Sandberg, J. (2011). Generating research questions through problematization. *Academy of Management Review, 36*(2), 247–271. https://doi.org/10.5465/amr.2009.0188

Alvesson, M., & Sandberg, J. (2013). *Constructing research questions. Developing interesting research through problematization*. SAGE Publications.

Alvesson, M., & Sandberg, J. (2014). Habitat and habitus: Boxed-in versus box-breaking research. *Organization Studies, 35*(7), 967–987. https://doi.org/10.1177/0170840614530916

Appenzeller, T. (2017). The AI revolution in science. *Science*. https://doi.org/10.1126/science.aan7064

Astley, W. G., & Zammuto, R. F. (1992). Organization science, managers, and language games. *Organization Science, 3*(4), 443–460. https://doi.org/10.1287/orsc.3.4.443

Australian Competition & Consumer Commission. (2011). Green marketing and the Australian Consumer Law. *Commonwealth of Australia*. https://www.accc.gov.au/system/files/Green%20marketing%20and%20the%20ACL.pdf. Accessed 29 December 2020.

Becker-Olsen, K., & Potucek, S. (2013). Greenwashing. In *Encyclopedia of Corporate Social Responsibility* (pp. 1318–1323). Springer. https://doi.org/10.1007/978-3-642-28036-8_104

Beder, S. (2002). Putting the boot in. *The Ecologist, 32*(3), 24–28.

Benkler, Y. (2019). Don't let industry write the rules for AI. *Nature, 569*(7755), 161–161.

Berrone, P., Fosfuri, A., & Gelabert, L. (2017). Does greenwashing pay off? Understanding the relationship between environmental actions and environmental legitimacy. *Journal of Business Ethics, 144*(2), 363–379. https://doi.org/10.1007/s10551-015-2816-9

Bietti, E. (2020). From ethics washing to ethics bashing: A view on tech ethics from within moral philosophy. In *Proceedings of ACM FAT* Conference (FAT* 2020)* (pp. 210–219). ACM. https://doi.org/10.1145/3351095.3372860

Bosse, D. A., & Phillips, R. A. (2016). Agency theory and bounded self-interest. *Academy of Management Review, 41*(2), 276–297. https://doi.org/10.5465/amr.2013.0420

Bowen, F. (2014). *After greenwashing: Symbolic corporate environmentalism and society*. Cambridge University Press. https://doi.org/10.1017/CBO9781139541213

Bromley, P., & Powell, W. W. (2012). From smoke and mirrors to walking the talk: Decoupling in the contemporary world. *Academy of Management Annals, 6*(1), 483–530. https://doi.org/10.5465/19416520.2012.684462

Brown, J. (2017). Why everyone is hating on IBM Watson—including the people who helped make it. *Gizmodo*. https://gizmodo.com/why-everyone-is-hating-on-watson-including-the-people-w-1797510888. Accessed 19 August 2020.

Bryson, J. J. (2020). The artificial intelligence of the ethics of artificial intelligence. In M. D. Dubber, F. Pasquale, & S. Das (Eds.), *The oxford handbook of ethics of AI* (pp. 3–25). Oxford University Press.

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. https://doi.org/10.1007/s10551-019-04226-4

Canadian Standards Association (CSA). (2008). Environmental claims: a guide for industry and advertisers. https://www.competitionbureau.gc.ca/eic/site/cb-bc.nsf/vwapj/guide-for-industry-and-advertisers-en.pdf/$FILE/guide-for-industry-and-advertisers-en.pdf. Accessed 29 December 2020

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*. https://doi.org/10.1177/1042258720934581

Chesney, R., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review, 107*, 1753–1820. https://doi.org/10.15779/Z38RV0D1

Christensen, L. T., Morsing, M., & Thyssen, O. (2017). License to critique: A communication perspective on sustainability standards. *Business Ethics Quarterly, 27*(2), 239–262. https://doi.org/10.1017/beq.2016.66

Christensen, L. T., Morsing, M., & Thyssen, O. (2020a). Talk-action dynamics: Modalities of aspirational talk. *Organization Studies*. https://doi.org/10.1177/0170840619896267

Christensen, L. T., Morsing, M., & Thyssen, O. (2020b). Timely hypocrisy? Hypocrisy temporalities in CSR communication. *Journal of Business Research, 114*, 327–335. https://doi.org/10.1016/j.jbusres.2019.07.020

Coeckelbergh, M. (2020). Artificial intelligence, responsibility attribution, and a relational justification of explainability. *Science and Engineering Ethics, 26*(4), 2051–2068. https://doi.org/10.1007/s11948-019-00146-8

Coeckelbergh, M. (2021). *Green leviathan or the poetics of political liberty: Navigating freedom in the age of climate change and artificial intelligence*. Routledge.

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling theory: A review and assessment. *Journal of Management, 37*(1), 39–67. https://doi.org/10.1177/0149206310388419

Corbin, K. (2020). Lawmaker wants to know why climate misinformation is rampant on youtube. *Forbes*. https://www.forbes.com/sites/kennethcorbin/2020/01/28/lawmaker-wants-to-know-why-climate-misinformation-is-rampant-on-youtube/?sh=3e8a7d7c3af9. Accessed 14 November 2020.

Cornelissen, J. P. (2005). Beyond compare: Metaphor in organization theory. *Academy of Management Review, 30*(4), 751–764. https://doi.org/10.5465/AMR.2005.18378876

Cornelissen, J. P., & Durand, R. (2014). Moving forward: Developing theoretical contributions in management studies. *Journal of Management Studies, 51*(6), 995–1022. https://doi.org/10.1111/joms.12078

Crawford, K., & Calo, R. (2016). There is a blind spot in AI research. *Nature, 538*(7625), 311–313. https://doi.org/10.1038/538311a

Dau-Schmidt, K. G. (2018). The impact of emerging information technologies on the employment relationship: New gigs for labor and employment law. *University of Chicago Legal Forum, 2017*(4), 63–94.

Delmas, M. A., & Burbano, V. C. (2011). The drivers of greenwashing. *California Management Review, 54*(1), 1–38. https://doi.org/10.1016/0737-6782(88)90039-2

den Hond, F., Rehbein, K. A., de Bakker, F. G. A., & Lankveld, H. K. (2014). Playing on two chessboards: reputation effects between corporate social responsibility (CSR) and corporate political activity (CPA). *Journal of Management Studies, 51*(5), 790–813. https://doi.org/10.1111/joms.12063

Department for Environment Food and Rural Affairs (DEFRA). (2010). Green Claims Guidance. *PB13453*. https://www.ukcpi.org/_Assets/custom-docs/publications/pb13453-green-claims-guidance.pdf. Accessed 18 January 2021.

DeVellis, R. F. (2017). *Scale development theory and applications* (4th ed.). SAGE Publications Inc.

Douek, E. (2019). Facebook's "Oversight Board:" Move fast with stable infrastructure and humility. *North Carolina Journal of Law and Technology, 21*(1), 1–78.

Du, X. (2014). How the market values greenwashing? Evidence from China. *Journal of Business Ethics, 128*(422), 547–574. https://doi.org/10.1007/s10551-014-2122-y

European Aviation Safety Agency. (2020). Artificial Intelligence Roadmap: A human-centric approach to AI in aviation. http://www.adfingo.com/easa-ai-roadmap-v10.pdf. Accessed 29 March 2020.

Federal Trade Commission (FTC). (2012). Part 260-guides for the use of environmental marketing claims. *77 FR 62124*. https://www.ftc.gov/sites/default/files/attachments/press-releases/ftc-issues-revised-green-guides/greenguides.pdf. Accessed 20 December 2020.

Fisher, B. (2019). Top 5 AI hires companies need to succeed in 2019. *KPMG*. https://info.kpmg.us/news-perspectives/technology-innovation/top-5-ai-hires-companies-need-to-succeed-in-2019.html. Accessed 31 July 2020.

Floridi, L. (2019). Translating principles into practices of digital ethics: Five risks of being unethical. *Philosophy and Technology, 32*(2), 185–193.

Futerra Sustainability Communications. (2009). The greenwash guide. https://www.silvaporto.com.br/wp-content/uploads/2017/09/GUIA_GREENWASHING.pdf.

Gatti, L., Seele, P., & Rademacher, L. (2019). Grey zone in—greenwash out: A review of greenwashing research and implications for the voluntary-mandatory transition of CSR. *International Journal of Corporate Social Responsibility, 4*(1), 1–15. https://doi.org/10.1186/s40991-019-0044-9

Gentner, D., & Smith, L. (2012). Analogical reasoning. In *Encyclopedia of human behavior* (2nd edn., Vol. 1, pp. 130–136). Elsevier. https://doi.org/10.1016/B978-0-12-375000-6.00022-7

Ginder, W., Kwon, W. S., & Byun, S. E. (2019). Effects of internal-external congruence-based CSR positioning: An attribution theory approach. *Journal of Business Ethics*. https://doi.org/10.1007/s10551-019-04282-w

Glozer, S., & Morsing, M. (2020). Helpful hypocrisy? Investigating 'double-talk' and irony in CSR marketing communications. *Journal of Business Research, 114*(2018), 363–375. https://doi.org/10.1016/j.jbusres.2019.08.048

Greenwood, R., Oliver, C., Lawrence, T. B., & Meyer, R. E. (Eds.). (2018). *The SAGE handbook of organizational institutionalism*. SAGE Publications Ltd.

Greenwood, R., Raynard, M., Kodeih, F., Micelotta, E. R., & Lounsbury, M. (2011). Institutional complexity and organizational responses. *The Academy of Management Annals, 5*(1), 317–371. https://doi.org/10.1080/19416520.2011.590299

Guo, R., Tao, L., Li, C. B., & Wang, T. (2017). A path analysis of greenwashing in a trust crisis among Chinese energy companies: the role of brand legitimacy and brand loyalty. *Journal of Business Ethics, 140*(3), 523–536. https://doi.org/10.1007/s10551-015-2672-7

Haack, P., Schoeneborn, D., & Wickert, C. (2012). Talking the talk, moral entrapment, creeping commitment? Exploring narrative dynamics in corporate responsibility standardization. *Organization Studies, 33*(5–6), 815–845. https://doi.org/10.1177/0170840612443630

Hadani, M., & Schuler, D. A. (2013). In search of El Dorado: The elusive financial returns on corporate political investments. *Strategic Management Journal, 34*(2), 165–181. https://doi.org/10.1002/smj.2006

Hagendorff, T. (2019). The ethics of ai ethics—an evaluation of guidelines. https://arxiv.org/abs/1903.03425.

Hagendorff, T., & Meding, K. (2020). The big picture: Ethical considerations and statistical analysis of industry involvement in machine learning research. http://arxiv.org/abs/2006.04541.

Hao, K. (2019). In 2020, let's stop AI ethics-washing and actually do something. *MIT Technology Review*. https://www.technologyreview.com/s/614992/ai-ethics-washing-time-to-act/. Accessed 22 January 2020.

Hao, K., & Stray, J. (2019). Can you make AI fairer than a judge? Play our courtroom algorithm game. *MIT Technology Review*. https://www.technologyreview.com/s/613508/ai-fairer-than-judge-criminal-risk-assessment-algorithm/. Accessed 5 February 2020.

Harvey, P., Madison, K., Martinko, M., Crook, T. R., & Crook, T. A. (2014). Attribution theory in the organizational sciences: The road traveled and the path ahead. *Academy of Management Perspectives, 28*(2), 128–146. https://doi.org/10.5465/amp.2012.0175

Hillman, A. J., Withers, M. C., & Collins, B. J. (2009). Resource dependence theory: A review. *Journal of Management, 35*(6), 1404–1427. https://doi.org/10.1177/0149206309343469

Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of Management, 21*(5), 967–988. https://doi.org/10.1177/014920639502100509

Huang, M.-H., & Rust, R. T. (2021a). A Framework for Collaborative Artificial Intelligence in Marketing. *Journal of Retailing*. https://doi.org/10.1016/j.jretai.2021.03.001

Huang, M.-H., & Rust, R. T. (2021b). Engaged to a robot? the Role of AI in service. *Journal of Service Research, 24*(1), 30–41. https://doi.org/10.1177/1094670520902266

IEEE Standards Association. (2020). The ethics certification program for autonomous and intelligent systems (ECPAIS). https://standards.ieee.org/industry-connections/ecpais.html. Accessed 20 March 2020.

Jeurissen, R. (1997). Integrating micro, meso and macro levels in business ethics. *Ethical Perspectives, 4*(4), 246–254. https://doi.org/10.2143/EP.4.4.562986

Jian, G., Shi, X., & Dalisay, F. (2014). Leader-member conversational quality: Scale development and validation through three studies. *Management Communication Quarterly, 28*(3), 375–403.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Johnson, D. G. (2015). Technology with no human responsibility? *Journal of Business Ethics, 127*(4), 707–715. https://doi.org/10.1007/s10551-014-2180-1

Johnson, K. (2019). How AI companies can avoid ethics washing. *VentureBeat*. https://venturebeat.com/2019/07/17/how-ai-companies-can-avoid-ethics-washing/. Accessed 18 January 2020.

Johnson, K. (2021). Google targets AI ethics lead Margaret Mitchell after firing Timnit Gebru. *VentureBeat*. https://venturebeat.com/2021/01/20/google-targets-ai-ethics-lead-margaret-mitchell-after-firing-timnit-gebru/. Accessed 25 January 2021.

Jones, D. R. (2012). Looking through the "greenwashing glass cage" of the green league table towards the sustainability challenge for UK universities. *Journal of Organizational Change Management, 25*(4), 630–647. https://doi.org/10.1108/09534811211239263

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Kalluri, P. (2020). Don't ask if artificial intelligence is good or fair, ask how it shifts power. *Nature, 583*(7815), 169–169. https://doi.org/10.1038/d41586-020-02003-2

Kamara, I. (2017). Co regulation in EU personal data protection: The case of technical standards and the privacy by design standardisation "mandate." *European Journal of Law and Technology, 8*(1), 1–24.

Ketokivi, M., Mantere, S., & Cornelissen, J. (2017). Reasoning by analogy and the progress of theory. *Academy of Management Review, 42*(4), 637–658. https://doi.org/10.5465/amr.2015.0322

Kinstler, L. (2020). Ethicists aim to save tech's soul. Will anyone let them? *Protocol LLC*. https://www.protocol.com/ethics-silicon-valley. Accessed 6 February 2020.

Knight, W. (2019). Google appoints an "AI council" to head o ff controversy, but it proves controversial. *MIT Technology Review*. https://www.technologyreview.com/2019/03/26/136376/google-appoints-an-ai-council-to-head-off-controversy-but-it-proves-controversial/. Accessed 5 January 2020.

Koene, A., Clifton, C., Hatada, Y., Webb, H., & Richardson, R. (2019). *A governance framework for algorithmic accountability and transparency*. (Panel for the Future of Science and Technology EPRS | European Parliamentary Research Service, Ed.). European Union.

Kopnina, H. (2019). Green-washing or best case practices? Using circular economy and Cradle to Cradle case studies in business education. *Journal of Cleaner Production, 219*, 613–621. https://doi.org/10.1016/j.jclepro.2019.02.005

Lange, D., & Washburn, N. T. (2012). Understanding attributions of corporate social irresponsibility. *Academy of Management Review, 37*(2), 300–326. https://doi.org/10.5465/amr.2010.0522

Laroche, M., Bergeron, J., & Barbaro-Forleo, G. (2001). Targeting consumers who are willing to pay more for environmentally friendly products. *Journal of Consumer Marketing, 18*(6), 503–520. https://doi.org/10.1108/EUM0000000006155

Laufer, W. S. (2003). Social accountability and corporate greenwashing. *Journal of Business Ethics, 43*(3), 253–261. https://doi.org/10.1023/A:1022962719299

Levin, S. (2019). Google scraps AI ethics council after backlash: "Back to the drawing board." *The Guardian*. https://www.theguardian.com/technology/2019/apr/04/google-ai-ethics-council-backlash. Accessed 22 January 2020.

Lim, S. J., & Phillips, J. (2008). Embedding CSR values: The global footwear industry's evolving governance structure. *Journal of Business Ethics, 81*(1), 143–156. https://doi.org/10.1007/s10551-007-9485-2

Long, B. S., & Driscoll, C. (2008). Codes of ethics and the pursuit of organizational legitimacy: Theoretical and empirical contributions. *Journal of Business Ethics, 77*(2), 173–189. https://doi.org/10.1007/s10551-006-9307-y

Lyon, T. P., & Montgomery, A. W. (2015). The means and end of greenwash. *Organization & Environment, 28*(2), 223–249. https://doi.org/10.1177/1086026615575332

Lyytinen, K., Majchrzak, A., & Song, M. (2017). Reinventing innovation management in a digital world. *MIS Quarterly: Management Information Systems, 41*(1), 223–238.

Marciniak, A. (2010). Greenwashing as an example of ecological marketing misleading practices. *Comparative Economic Research, 12*(1–2), 49–59.

Marquis, C., Toffel, M. W., & Zhou, Y. (2016). Scrutiny, norms, and selective disclosure: a global study of greenwashing. *Organization Science, 27*(2), 483–504. https://doi.org/10.1287/orsc.2015.1039

Martin, K. (2019). Ethical implications and accountability of algorithms. *Journal of Business Ethics, 160*(4), 835–850. https://doi.org/10.1007/s10551-018-3921-3

Martin, K., Shilton, K., & Smith, J. (2019). Business and the ethical implications of technology: introduction to the symposium. *Journal of Business Ethics, 160*(2), 307–317. https://doi.org/10.1007/s10551-019-04213-9

Matejek, S., & Gössling, T. (2014). Beyond legitimacy: A case study in BP's "Green Lashing." *Journal of Business Ethics, 120*(4), 571–584. https://doi.org/10.1007/s10551-013-2006-6

McLennan, S., Fiske, A., Celi, L. A., Müller, R., Harder, J., Ritt, K., et al. (2020). An embedded ethics approach for AI development. *Nature Machine Intelligence, 2*(9), 488–490. https://doi.org/10.1038/s42256-020-0214-1

McMillan, D., & Brown, B. (2019). Against ethical AI. In *Proceedings of the Halfway to the Future Symposium 2019* (pp. 1–3). ACM. https://doi.org/10.1145/3363384.3363393

Melé, D., & Armengou, J. (2016). Moral legitimacy in controversial projects and its relationship with social license to operate: A case study. *Journal of Business Ethics, 136*(4), 729–742. https://doi.org/10.1007/s10551-015-2866-z

Mellahi, K., Frynas, J. G., Sun, P., & Siegel, D. (2016). A review of the nonmarket strategy literature. *Journal of Management, 42*(1), 143–173. https://doi.org/10.1177/0149206315617241

Merriam-Webster Dictionary. (2020). Greenwashing. *Merriam-Webster Dictionary Web Site*. https://www.merriam-webster.com/dictionary/greenwashing. Accessed 5 August 2020.

Metz, C. (2019). A.I. is learning from humans. Many humans. *The New York Times*. Artificial intelligence is being taught by thousands of office workers around the world. It is not exactly futuristic work. Accessed 15 January 2020.

Metz, C., & Wakabayashi, D. (2020). Google researcher says she was fired over paper highlighting bias in A.I. *The New York Times*. https://www.nytimes.com/2020/12/03/technology/google-researcher-timnit-gebru.html. Accessed 4 December 2020.

Metzinger, T. (2019). EU guidelines: Ethics washing made in Europe. *Der Tagesspiegel*. https://www.tagesspiegel.de/politik/eu-guidelines-ethics-washing-made-in-europe/24195496.html. Accessed 25 November 2019.

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence, 1*(11), 501–507.

Mittelstadt, B. D., Stahl, B. C., & Fairweather, N. B. (2015). How to shape a better future? Epistemic difficulties for ethical assessment and anticipatory governance of emerging technologies. *Ethical Theory and Moral Practice, 18*(5), 1027–1047. https://doi.org/10.1007/s10677-015-9582-8

Mozafari, N., Weiger, W., & Hammerschmidt, M. (2020). Resolving the chatbot disclosure dilemma: Leveraging selective self-presentation to mitigate the negative effect of chatbot disclosure. In *Proceedings of the Hawaii International Conference on System Sciences* (pp. 2916–2923). https://www.researchgate.net/profile/Maik_Hammerschmidt/publication/344850372_Resolving_the_Chatbot_Disclosure_Dilemma_Leveraging_Selective_Self-Presentation_to_Mitigate_the_Negative_Effect_of_Chatbot_Disclosure/links/5f940c06299bf1b53e4088b6/Resolving-th.

Murray, A., Skene, K., & Haynes, K. (2017). The circular economy: An interdisciplinary exploration of the concept and application in a global context. *Journal of Business Ethics, 140*(3), 369–380. https://doi.org/10.1007/s10551-015-2693-2

Nersessian, N. J. (2008). *Creating scientific concepts*. The MIT Press.

Ng, A. (2016). What AI can and can't do right now. *Harvard Business Review*, pp. 2–4. https://hbr.org/2016/11/what-artificial-intelligence-can-and-cant-do-right-now.

Noto La Diega, G. (2018). Against the dehumanisation of decision-making. Algorithmic decisions at the crossroads of intellectual property, data protection, and freedom of information. *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*. https://doi.org/10.31228/osf.io/s2jnk

Nyilasy, G., Gangadharbatla, H., & Paladino, A. (2014). Perceived greenwashing: The interactive effects of green advertising and corporate environmental performance on consumer reactions. *Journal of Business Ethics, 125*(4), 693–707. https://doi.org/10.1007/s10551-013-1944-3

Obar, J. A., & Oeldorf-Hirsch, A. (2020). The biggest lie on the Internet: Ignoring the privacy policies and terms of service policies of social networking services. *Information, Communication & Society, 23*(1), 128–147. https://doi.org/10.1080/1369118X.2018.1486870

Obradovich, B. N., Powers, W., Cebrian, M., Rahwan, I., & Content, R. (2019). Beware corporate "machinewashing" of AI. *Media MIT*. https://www.media.mit.edu/articles/beware-corporate-machinewashing-of-ai/. Accessed 15 March 2020.

Ochigame, R. (2019). The invention of "Ethical AI": How big tech manipulates academia to avoid regulation. *The Intercept*. https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/. Accessed 22 January 2020.

Ochigame, R., Lundgard, A., Dominguez, A. E., Zong, J., Ballard, G., Haslanger, S., et al. (2019). The struggle to democratize MIT Undemocratic committees won't stop unethical financial partnerships. *The Tech*. https://thetech.com/2019/10/23/struggle-democratize-mit. Accessed 28 January 2020.

Orange, E. (2010). From eco-friendly to eco-intelligent. *Futurist, 44*(5), 29–32.

Oxford English Dictionary. (2012). Greenwashing, n. http://www.oed.com/view/Entry/249122. Accessed 28 February 2018.

Palazzo, G., & Scherer, A. G. (2006). Corporate legitimacy as deliberation: A communicative framework. *Journal of Business Ethics, 66*(1), 71–88.

Papazoglou, A. (2019). Silicon valley' s secret philosophers should share their work opinion. *Wired*. https://www.wired.com/story/silicon-valleys-secret-philosophers-should-share-their-work/. Accessed 14 January 2020.

Petrenko, O. V., Aime, F., Ridge, J., & Hill, A. (2016). Corporate social responsibility or CEO narcissism? CSR motivations and organizational performance. *Strategic Management Journal, 37*(2), 262–279. https://doi.org/10.1002/smj.2348

Pizzetti, M., Gatti, L., & Seele, P. (2019). Firms talk, suppliers walk: Analyzing the locus of greenwashing in the blame game and introducing 'Vicarious Greenwashing.' *Journal of Business Ethics*. https://doi.org/10.1007/s10551-019-04406-2

Pope, S., & Wæraas, A. (2016). CSR-washing is rare: A conceptual framework, literature review, and critique. *Journal of Business Ethics, 137*(1), 173–193. https://doi.org/10.1007/s10551-015-2546-z

Rakova, B., Yang, J., Cramer, H., & Chowdhury, R. (2020). Where responsible AI meets reality: Practitioner perspectives on enablers for shifting organizational practices. *Proceedings of the ACM on Human-Computer Interaction, 1*(1), 1–18.

Ramdhony, D. (2018). The implications of mandatory corporate social responsibility: A literature review perspective. *Theoretical Economics Letters, 08*(03), 432–447. https://doi.org/10.4236/tel.2018.83031

Rehbein, K., den Hond, F., & Bakker, F. G. A. (2018). Aligning adverse activities? Corporate social responsibility and political activity (pp. 295–324). Emerald Publishing Limited. https://doi.org/10.1108/S2514-175920180000002008

Rességuier, A., & Rodrigues, R. (2020). AI ethics should not remain toothless! A call to bring back the teeth of ethics. *Big Data & Society, 7*(2), 1–5. https://doi.org/10.1177/2053951720942541

Roose, K. (2019). The hidden automation agenda of the davos elite: the New York Times. *The New York Times*. https://www.nytimes.com/2019/01/25/technology/automation-davos-world-economic-forum.html. Accessed 15 July 2020.

Rust, R. T., & Huang, M.-H. (2021). Moral, ethical, and governance implications. In *The Feeling Economy* (pp. 129–138). Springer International Publishing. https://doi.org/10.1007/978-3-030-52977-2_12

Satariano, A. (2020). Silicon valley heads to europe, nervous about new rules. *The New York Times*. https://www.nytimes.com/2020/02/16/technology/europe-new-AI-tech-regulations.html?action=click&module=Well&pgtype=Homepage&section=Technology.

Scholz, M., & de los Reyes, G., & Smith, N. C. (2019). The enduring potential of justified hypernorms. *Business Ethics Quarterly, 29*(03), 317–342. https://doi.org/10.1017/beq.2018.42

Scott, W. R. (2014). *Institutions and organizations: Ideas, interests, and identities* (4th ed.). SAGE Publications Inc.

Seele, P., Dierksmeier, C., Hofstetter, R., & Schultz, M. D. (2021). Mapping the ethicality of algorithmic pricing: A review of dynamic and personalized pricing. *Journal of Business Ethics, 170*(4), 697–719. https://doi.org/10.1007/s10551-019-04371-w

Seele, P., & Gatti, L. (2017). Greenwashing revisited: In search of a typology and accusation-based definition incorporating legitimacy strategies. *Business Strategy and the Environment, 26*(2), 239–252.

Seele, P., & Lock, I. (2015). Instrumental and/or deliberative? A typology of CSR communication tools. *Journal of Business Ethics, 131*(2), 401–414. https://doi.org/10.1007/s10551-014-2282-9

Sharkey, N. (2018). Mama mia it's sophia: a show robot or dangerous platform to mislead? *Forbes*. https://www.forbes.com/sites/noelsharkey/2018/11/17/mama-mia-its-sophia-a-show-robot-or-dangerous-platform-to-mislead/?sh=7648eb477ac9. Accessed 5 December 2020.

Sheehy, B. (2014). Defining CSR: problems and solutions. *Journal of Business Ethics, 131*(3), 625–648. https://doi.org/10.1007/s10551-014-2281-x

Shirodkar, V., Beddewela, E., & Richter, U. H. (2018). Firm-level determinants of political CSR in emerging economies: Evidence from India. *Journal of Business Ethics, 148*(3), 673–688. https://doi.org/10.1007/s10551-016-3022-0

Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review, 20*(3), 571–610. https://doi.org/10.5465/AMR.1995.9508080331

Suddaby, R., Bitektine, A., & Haack, P. (2017). Legitimacy. *Academy of Management Annals, 11*(1), 451–478. https://doi.org/10.5465/annals.2015.0101

Susser, D. (2019). Ethics alone can't fix big tech. *Slate*. https://slate.com/technology/2019/04/ethics-board-google-ai.html. Accessed 22 January 2020.

Swedberg, R. (Ed.). (2014). *Theorizing in Social science*. Stanford University Press.

Szabo, S., & Webster, J. (2020). Perceived greenwashing: The effects of green marketing on environmental and product perceptions. *Journal of Business Ethics*. https://doi.org/10.1007/s10551-020-04461-0

TerraChoice. (2010). The sins of greenwashing home and family edition 2010. *Underwriters Laboratories*. http://sinsofgreenwashing.org/findings/greenwashing-report-2010/. Accessed 12 April 2020.

Theodorou, A., & Dignum, V. (2020). Towards ethical and socio-legal governance in AI. *Nature Machine Intelligence, 2*(1), 10–12. https://doi.org/10.1038/s42256-019-0136-y

Truby, J. (2020). Governing artificial intelligence to benefit the UN sustainable development goals. *Sustainable Development, 2*, 2048.

Umbrello, S., & van de Poel, I. (2020). Mapping value sensitive design onto AI for social good principles. *Online Document*. https://philpapers.org/archive/UMBMVS.pdf. Accessed 4 August 2020.

Vaughan, D. (2014). 3. Analogy, cases, and comparative social organization. In *Theorizing in social science* (pp. 61–84). Stanford University Press. https://doi.org/10.1515/9780804791199-005

Verbeke, A., & Greidanus, N. (2009). The end of the opportunism vs trust debate: Bounded reliability as a new envelope concept in research on MNE governance. *Journal of International Business, 40*(9), 1471–1495.

Vincent, J. (2021). Google is poisoning its reputation with AI researchers. *The Verge*. https://www.theverge.com/2021/4/13/22370158/google-ai-ethics-timnit-gebru-margaret-mitchell-firing-reputation. Accessed 16 April 2021.

Waddell, K. (2019). The dangers of "AI washing." *Axios*. https://www.axios.com/ai-washing-hidden-people-00ab65c0-ea2a-4034-bd82-4b747567cba7.html. Accessed 16 January 2020.

Wagner, B. (2018). Ethics as an escape from regulation: From ethics-washing to ethics-shopping? In B. Emre, B. Irina, J. Liisa, & H. Mireille (Eds.), *Being profiled - cogitas ergo sum* (pp. 84–89). Amsterdam University Press.

Wagner, B., & Winkler, T. (2019). Comparing routing predictions: travel time estimates and user accountability in navigation apps. In *Twenty-Seventh European Conference on Information Systems (ECIS2019)* (pp. 1–8).

Wakabayashi, D. (2020). Big tech funds a think tank pushing for fewer rules. For Big Tech. *The New York Times*. https://www.nytimes.com/2020/07/24/technology/global-antitrust-institute-google-amazon-qualcomm.html. Accessed 27 July 2020.

Walker, K., & Wan, F. (2012). The harm of symbolic actions and green-washing: Corporate actions and communications on environmental performance and their financial implications. *Journal of Business Ethics, 109*(2), 227–242. https://doi.org/10.1007/s10551-011-1122-4

Walter, B. L. (2010). *Verantwortliche Unternehmensführung überzeugend kommunizieren: Strategien für mehr Transparenz und Glaubwürdigkeit*. Gabler Verlag.

Whelan, J., & Demangeot, C. (2015). Signaling theory. In *Wiley encyclopedia of management* (pp. 1–1). Wiley https://doi.org/10.1002/9781118785317.weom090243

Williamson, O. E. (1971). The vertical integration of production: market failure considerations. *The American Economic Review, 61*(2), 112–123.

Winkler, P., Etter, M., & Castelló, I. (2020). Vicious and virtuous circles of aspirational talk: From self-persuasive to agonistic CSR rhetoric. *Business and Society, 59*(1), 98–128. https://doi.org/10.1177/0007650319825758

Wright, J. D., Dorsey, E., Rybnicek, J., & Klick, J. (2018). Requiem for a paradox: The dubious rise and inevitable fall of hipster antitrust. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3249524

Wu, Y., Zhang, K., & Xie, J. (2020). Bad greenwashing, good greenwashing: Corporate social responsibility and information transparency. *Management Science, 66*(7), 3095–3112. https://doi.org/10.1287/mnsc.2019.3340

Yeung, K. (2019). Responsibility and AI. *Council of Europe DGI(2019)05*. https://rm.coe.int/responsability-and-ai-en/168097d9c5. Accessed 15 February 2020.

Yeung, K., Howes, A., & Pogrebna, G. (2020). AI governance by human rights-centred design, deliberation and oversight: An end to ethics washing. In M. Dubber, F. Pasquale, & S. Das (Eds.), *Oxford handbook of the ethics of artificial intelligence.* Oxford University Press.

Yuste, R., Goering, S., Arcas, B. A., Bi, G., Carmena, J. M., Carter, A., et al. (2017). Four ethical priorities for neurotechnologies and AI. *Nature, 551*(7679), 159–163. https://doi.org/10.1038/551159a

Zanasi, C., Rota, C., Trerè, S., & Falciatori, S. (2017). An assessment of the food companies sustainability policies through a greenwashing indicator. *International Journal on Food System Dynamics, 74*, 61–81. https://doi.org/10.18461/pfsd.2017.1707

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The fight for a human future at the new frontier of power*. Profile Books Ltd.

Zuboff, S. (2021). The coup we are not talking about. *The New York Times*. https://www.nytimes.com/2021/01/29/opinion/sunday/facebook-surveillance-society-technology.html. Accessed 29 January 2021.