Journal of Business Ethics (2022) 178:895–916
https://doi.org/10.1007/s10551-022-05050-z

ORIGINAL PAPER

The Dawn of the AI Robots: Towards a New Framework of AI Robot Accountability

Zsófia Tóth[1] · Robert Caruana[2] · Thorsten Gruber[3] · Claudia Loebbecke[4]

Received: 5 October 2020 / Accepted: 25 January 2022 / Published online: 2 March 2022

Abstract
Business, management, and business ethics literature pay little attention to the topic of AI robots. The broad spectrum of potential ethical issues pertains to using driverless cars, AI robots in care homes, and in the military, such as Lethal Autonomous Weapon Systems. However, there is a scarcity of in-depth theoretical, methodological, or empirical studies that address these ethical issues, for instance, the impact of morality and where accountability resides in AI robots' use. To address this dearth, this study offers a conceptual framework that interpretively develops the ethical implications of AI robot applications, drawing on descriptive and normative ethical theory. The new framework elaborates on how the locus of morality (human to AI agency) and moral intensity combine within context-specific AI robot applications, and how this might influence accountability thinking. Our theorization indicates that in situations of escalating AI agency and situational moral intensity, accountability is widely dispersed between actors and institutions. 'Accountability clusters' are outlined to illustrate interrelationships between the locus of morality, moral intensity, and accountability and how these invoke different categorical responses: (i) illegal, (ii) immoral, (iii) permissible, and (iv) supererogatory pertaining to using AI robots. These enable discussion of the ethical implications of using AI robots, and associated accountability challenges for a constellation of actors—from designer, individual/organizational users to the normative and regulative approaches of industrial/governmental bodies and intergovernmental regimes.

Keywords Artificial intelligence · AI robots · Descriptive ethical theory · Normative ethical theory · Locus of morality · Moral intensity · Accountability

Introduction

Artificial intelligence (AI) robots are intelligent, semi-autonomous machines, software and systems that have the increasing ability to formulate decisions in collaboration with humans or on their own, to support humans. They enable higher efficiency and increase quality of life (Čaić et al., 2018) and so can have a profound effect on business and society. AI robots assist elderly care (Broekens et al., 2009; Jiang & Cameron, 2020), support medical diagnoses (Yoon & Lee, 2019), and ease transportation in the case of autonomous cars (Hassan et al., 2018). At the same time, however, using AI robots triggers societal changes and thereby yields ethical implications (Alles & Gray, 2020; Veruggio et al., 2016; Wirtz et al., 2018). As Westerlund (2020) suggests, AI robots may come to significantly shape the socio-political order over time, raising ethical issues and accountability concerns at the highest level, as is already the case with algorithms and personal data harvesting (e.g., the Cambridge Analytica scandal, Wang et al., 2020).

Although it is difficult to predict the pace of technological innovation in AI robots, a business and management-based discussion of AI robot ethics is necessary to mitigate future risks (Russell et al., 2015), to assist both the codifying of AI robot behavior (Gunkel, 2012) as well as developing accountability mechanisms in business settings. Although the field of *business* ethics is relatively new within the ethics domain, it can help guide individual and organizational users alike, who are aiming to better manage their own, and

✉ Zsófia Tóth
zsofia.toth@durham.ac.uk

1 Business School, Durham University, Durham, UK

2 Business School, University of Nottingham, Nottingham, UK

3 Business School, University of Loughborough, Loughborough, UK

4 University of Cologne, Cologne, Germany

in case of organizations, their employees' ethical behaviors (Trevino & Brown, 2004). While some users already may "have a higher level of global awareness" to act ethically (Huang & Rust, 2011, p. 44), several others need more guidance. In this paper we summarize insights from normative and descriptive ethical theory, drawing on the former's capacity to determine actionable, categorical responses to ethical challenges of AI robots (of which we identify four categories), and the latter's *explanatory* capacity in relation to ethical dimensions shaping AI robot contexts (of which we cite two—moral agency and intensity). These theoretical elements combine to provoke thinking about 'accountability clusters' (multitude of actors, levels, and institutions) needed to govern AI robot applications in business.

To our best knowledge, applying normative and descriptive ethics to AI robots in business settings is a novel approach, as it enables one to integrate concepts from yet unrelated knowledge domains with concepts pertaining to AI robots and the reflections on their practical matters. We use these insights to develop a new framework that incorporates the following constructs: locus of morality (human to AI agency), moral intensity and accountability dispersal, accountability clusters, and the four ethical categories of illegal, immoral, permissible, and supererogatory. Whereas the locus of morality depends on where moral decision-making lies (Kagan, 2018), moral intensity refers to the extent of issue-related moral imperative across different situations and considers the impact a single action can have on multiple victims or beneficiaries (Jones, 1991). However, we are also mindful that ethical implications of AI robots stem from a unique web of interrelationships between loosely connected actors such as AI robot designers, individual and organizational users, industry and government bodies as well as civil society groups. Therefore, one must consider new forms of AI robot accountability—which we describe theoretically as 'accountability clusters.' These are the networks of relevant actors positioned at different levels who constitute mechanisms for AI robot accountability (e.g., personal/professional, organizational, institutional, supra-territorial).

Regarding this important phenomenon, we broadly adhere to Beu and Buckley's (2001, p. 65) definition: "Accountability is the perceived need to defend or justify behaviors to an audience with reward/sanction authority, where the rewards/sanctions are perceived to be contingent upon the audience evaluation of such conduct." However, as Buhmann et al. (2019) note with regard to 'Algorithmic Accountability,' there may be special, discrete accountability characteristics specific to AI and system learning technologies (such as technical and strategic opacity) that render expectations of accountability as highly fluid. AI robot applications have the potential to significantly complicate traditional, formal mechanisms of accountability (e.g., workplace tribunal) by dispersing moral agency between potentially numerous agents, who may be responsible for a grievous harm, but whom are spatially, organizationally, and even temporally disconnected. Accountability dispersal, then, is the extent to which accountability spreads across different actors and levels, where high accountability dispersal poses communication and coordination challenges stakeholders face when it comes to ensuring the ethical use of AI robots. Our theorization of accountability dispersal recommends investigating inter-linkages between actors within and across levels, echoing the rich tradition of scholarly multilevel research into related topics like governance and Corporate Social Responsibility (Balakrishnan et al., 2017; François et al., 2019; Young & Marais, 2012). Accordingly, while we broadly follow a micro–meso–macro analytical pathway, our four 'accountability clusters' more precisely detail the nature of actors and characteristics of contexts surrounding AI robot applications. Therefore, our aim in this Special Issue of the *Journal of Business Ethics* is to promote the building of 'theoretical bridges' (Hitt et al., 2007) across levels within future business and society research into AI and AI robots. We explore—via four clusters—how different AI application contexts interact with notions of moral intensity, agency, and accountability, extracting clusters of variously positioned actors, and prompting consideration of some fundamental outcomes for accountability of AI robots' use. Therefore, we highlight the value of applying descriptive and normative theory to a vital business–society issue.

The potential contributions of developing a new conceptual framework for AI robot accountability are as follows: first, to expand the current understanding of the ethical implications of AI robot applications, we formulate the aforementioned accountability clusters. These clusters indicate necessary actors and activities to ensure accountability, for instance, corporations that design and implement AI robots, industry, governments/regulators, and civil society organizations. Here, we define 'accountability clusters' as a nexus of relevant actors positioned at different levels that constitute mechanisms for AI robot accountability (e.g., individual, organizational, industrial/governmental, supra-territorial), which would serve to govern un/intended ethical transgressions in AI robot application contexts.

Second, for each accountability cluster, we integrate four normative ethical categories of illegal, immoral, and permissible from Heath (2014), who outlined them as pillars of the "market failures approach to business ethics" that is part of normative business ethics. In addition, we extend Heath's work by incorporating the category of supererogatory use (Driver, 1992), characterized by creating excess value, to also cover potential positive implications of using AI robots. The concept of supererogatory use also is rooted in normative business ethics. Incorporating both negative (i.e., illegal, immoral) and neutral/positive (i.e., permissible, supererogatory)

aspects opens avenues for the ethical investigation of innovation around AI robots that enables one to create a more nuanced understanding compared to the binary view of something that is either ethical or unethical (e.g., Bommer et al., 1987; Constantinescu & Kaptein, 2015; Khalil, 1993). This translates into moving from 'yes or no' ethical evaluations towards the quadrangle of 'yes, please' (supererogatory)/'alright' (permissible)/'rather not' (immoral)/'not at all' (illegal). Further, these categories represent different layers of moral intensity with supererogatory use as the least morally intense. We derive our framework from drawing on examples of AI robot uses, with a special emphasis on human–technology interaction from an ethical and regulatory perspective (Johnson, 2015; Lobschat et al., 2021; Wirtz et al., 2018).

Third, we address the scarcity of—especially the macro-level—studies on business ethics that use AI robots (with Wirtz et al., 2018 as an exception) to extend the literature in the field of normative business ethics. Finally, our approach can inform policy recommendations for regulatory bodies, firms, and individuals regarding developing and controlling ethically astute AI robots, for instance through using the framework for scenario planning, stakeholder mapping, and AI robot design. Drawing on the proposed new business ethics framework, we offer some consideration points for regulatory intervention related to the AI robots' learning behaviors and their role in decision-making processes.

The study is structured as follows: we first seek conceptual clarity pertaining to AI robots. Then we discuss ethics, with special regard to business ethics that serves as the basis for the new framework outlined in the subsequent section, along with accountability clusters. Next, we outline ethical implications in the discussion section that includes regulatory implications. Finally, we highlight limitations and directions for future research, including research topics and research questions. The first author is grateful to Dr Sareh Pouryousefi for the helpful discussions on business ethics.

AI, Robots, and AI Robots

Conceptual clarity is required to evaluate relevant ethical implications and so the section below focuses on providing definitions of AI, robots, as well as AI robots. The three are connected phenomena and the third phenomenon, AI robots, is of special interest for this study. While we focus on AI robots, one also can apply the outlined ethical considerations to a wider range of AI applications.

AI

AI refers to developing intelligent, autonomous systems that can perform tasks otherwise attributed to human intelligence, such as visual or speech recognition, language translation, and reacting to events in the environment (King, 2017). Thus, AI constitutes of intelligent software. Algorithms—and especially decision-making algorithms—based on machine learning techniques are inherent parts of AI (Martin, 2019). AI is commonly associated with machine programing to enable participation in human-like thought processes such as learning, reasoning, and self-correction (Benlian et al., 2019). AI spreads across a variety of activity areas such as machine learning, knowledge representation, modeling human cognition, data science, augmented reality, computer imaging, audio–visual signaling, and natural language processing, just to mention a few. The outputs AI generates include information, human–computer communication, and even physical objects (Baskerville et al., 2020). AI also increasingly supports, and in some areas even substitutes, human decision-making (Baskerville et al., 2020).

Robots

Historically, the concept of robots referred to automatic devices that perform functions ordinarily ascribed to human intelligence (Calo, 2017). Robots act upon codifiable, predetermined goals and follow cognitive structures to adapt to their environment. They can recognize part of their environment such as physical objects or human voices (Aleksander, 2017) and carry out specific, pre-programmed actions, for instance moving objects and interacting with humans (Admoni & Scassellati, 2017). A robot's level of control is limited: humans have permission to correct or stop a robot's actions (Zieba et al., 2011). With their information processing capacity and domain-specific cognitive abilities, robots often exceed human performance in various areas (Ma & McGroarty, 2017). Further specifications regard robots as automatically controlled, reprogrammable, multipurpose machines, which one can either fix in place or make mobile for use (c.f. the ISO 8373 standard). Traditionally, complex IT systems that enable learning and support decision-making did not support robots. While several robots match these characteristics, the traditional approach focuses more on the physical entity of robots.

AI Robots

Advancing the definition of AI robots and continuously adjusting it to the latest level of technological development have been ongoing challenges. Among others, European legislation in the form of the Parliament's resolution to the Commission on Civil Law Rules on Robotics (EP

2015/2103, INL) has called for an up-to-date, specific, and actionable definition that encompasses both AI and robots. For this paper, *the working definition of AI robots is that they are semi-autonomous, insensate entities that exhibit behaviors of living beings and possess the abilities of learning and decision-making to facilitate human activity* (based on Aleksander, 2017; King, 2017). This working definition draws on the definition of robots, with special regard to performing activities through sensing and adapting to the environment (Aleksander, 2017), as well as on the definition of AI, especially pertaining to human-like intelligence but not feelings (King, 2017). Some AI robots can have a physical representation, thus are machines with intelligent software (e.g., Nao or Pepper), while others are only virtually represented without the necessary physical representation of a specific machine (e.g., Siri or Alexa). Consequently, following Wirtz et al. (2018), scholars regard virtual AI software with the ability to learn over time and the capacity for autonomous action as an example of AI robots.

AI robots combine automation mechanisms and sophisticated learning and decision-making abilities to support humans. However, AI robots are still incapable of processing or expressing emotions and other vital aspects of human-to-human communication (Ciborra & Willcocks, 2006). Partly due to the lack of empathy, AI robots' learning processes differ from human learning (Kamishima et al., 2018) in that AI robots require shortened learning cycles (Bilgeri et al., 2019), yielding a higher capacity for processing large amounts of information (Bera et al., 2019) and thus can reduce human workload. AI robots typically consist of an agent (that can be a physical entity such as an AI robot but also a software) and its environment (in which the agent acts and has an intelligent connection to, for instance, through sensors; the environment of an agent may contain further agents) (Choudhary et al., 2016). Although AI robots already appeared as witnesses in front of the court, they are typically, though perhaps not accurately, regarded as morally passive tools (Westerlund, 2020). Consequently, the legal standing of AI robots and their liability are still under discussion (Calo, 2017) but they can assist judges' work when it comes to preparing background materials and assessing expert testimonies (Katz, 2013). AI robots' decisions are becoming increasingly important and part of daily life. However, it has been an ongoing challenge to distinguish between AI robots' decisions and the ones exclusively suitable for humans (Baskerville et al., 2020; Ciborra & Willcocks, 2006).

AI Robots and Ethics

Literature on AI robots identifies concentrated accountability around users (e.g., Buhmann et al., 2019; Westerlund, 2020), manufacturers (Bench-Capon, 2020; Buhmann et al., 2019), other organizations such as governments (Wright & Schultz, 2018) and partially AI robots themselves (Bench-Capon, 2020). Researchers have applied a variety of approaches from virtue ethics (Bench-Capon, 2020) to social contract theory (Wright & Schultz, 2018). The focus of these studies varies across several topics, between the abdication of human responsibility (Allen & Wallach, 2014) to algorithmic accountability (Buhmann et al., 2019). A common pattern is that authors characteristically point to concentrated accountability and open the discussion on the need for an ethical dimension in the context of AI robots. Table 1 presents some key points from the literature and position our study against the identified sources.

Focus on Normative Business Ethics

Ethics refer to the implicit and explicit norms and principles one should follow in the absence of governmental guidelines or other external regulatory regimes (Heath, 2008). Interdisciplinary ethical analysis incorporates a wide range of managerial, economic, social, technical, and legal issues (Zsolnai, 2006) and discusses regulatory actions for mitigating ethical concerns. Business ethics is an interdisciplinary field that pertains to a range of normative issues in markets, including questions surrounding individual behavior and responsibility, organizational and institutional ethics, as well as the just design of markets, regulations, and political oversight (Norman, 2011). Normative business ethics refers to a field of business ethics that investigates how ethics can inform decision-making (Hasnas, 1998). Kagan (2018) suggests that normative business ethics involves substantive proposals about how to act, how to live, how to do business, and what kind of person to be. It identifies morally acceptable actions under given conditions and derives key regulatory/management protocols (Cropanzano et al., 2013). As opposed to normative ethics, meta-ethics (Miller, 2003) aim at delineating moral concepts and justifying moral theory without suggesting what comprises 'right' or 'wrong.' However, normative business ethics studies moral principles and develops guidance for resolving individual/institutional moral dilemmas and market design. Furthermore, normative business ethics raises questions about how to *engage* with non-human entities.

Within the field of normative business ethics, the literature distinguishes between *virtue ethics* (Hursthouse, 1999), which is primarily concerned with evaluating an individual's inner states and the fit between actions and the character. *Deontology* (Alexander & Moore, 2007) is the study of duty, i.e., what moral obligation requires us to do. *Consequentialism* (Peterson, 2013) argues that the act's consequences determine an act's moral rightness. The *ethical economy approach* (Koslowski, 2001) argues for combining ethics

Table 1 Some key points from the literature on AI robots and the positioning of this study

Studies	Concept	Focus	Approach/framework	Actors with accountability			
				User	Manufacturer	AI robot	Other
Allen and Wallach (2014) (in Lin et al., eds.)	The development of Artificial Moral Agents (AMAs)	Abdication of human responsibility	AMA development is assessed on the Autonomy and Ethical sensitivity axes: from today's low autonomy and low ethical sensitivity an increase is envisioned on both sides through operational morality, to functional morality and potentially further towards full moral agency of AI robots	n/a	Yes, implicitly (attention is raised on the illusion that there is a technological fix to AI-related risks)	Partially (increasingly but with the limitation that full moral agency is unachievable and AI robots need a moral code that is different from human morality)	n/a
Bench-Capon (2020)	Ethical agents and their implementation within the virtue ethics approach; ethical agent = an agent that behaves in a way that would be considered ethical in a human being	How core ethical approaches: deontology, consequentialism, and virtue ethics relate to the implementation of ethical agents	Virtue ethics (there are morally good/bad actions and motives; morally good actions exemplify virtues and morally bad actions exemplify vices). In case of humans, being selfish is unethical, being altruistic is ethica,l and being sacrificial is supererogatory	n/a	Yes (the designer can constrain the agent)	Partially (depending on the ethical approach; an agent can potentially generate itself and be influenced by other agents; software agents do not have needs, or if they do, not these needs)	Potentially (the potential interconnectedness between agents is acknowledged in Multi-Agent Systems)

Table 1 (continued)

Studies	Concept	Focus	Approach/framework	Actors with accountability			
				User	Manufacturer	AI robot	Other
Buhmann et al. (2019)	Algorithmic accountability: algorithms come with an accountability issue that is relatively recent but increasingly a public concern. Self-learning algorithms as obstacles	How to manage algorithmic accountability; reputational concerns, engagement strategies, and discourse disciplines are assessed	"Reverse engineering" is identified as a potential strategy, i.e., to study the subject by observing inputs and outputs rather than trying to understand flight by studying feathers. This suggests a discourse-ethical approach. Manipulative, adaptive, and moral approaches are identified for organizations in relation to algorithmic accountability to manage engagement strategies	Yes organizational users (businesses)	Yes developers	No	Indirectly: civil society organizations are central to make complex technical issues accessible and understandable to a wider public
Martin (2019)	Algorithmic mistakes and the role of algorithms in decision-making	Types of mistakes (Type 1 and 2 errors) and their social embeddedness are studied	Algorithm-related decision accountability is assessed along the axes of "degree of reflection" (~ability to correct mistakes) and "social embeddedness" (~the extent to which users can identify mistakes)	Yes (user's responsibility is high when their abilities to identify and fix mistakes are high)	Yes (developer: developer's responsibility is high when the user's abilities to identify and fix mistakes are low)	No	Implicitly, social context (+manufacturing example: assign someone to correct the mistake of a machine-generated decision)
Torresen (2018)	Similarity of AI robots to humans and its effects on human–AI robot interaction; risk management	Future jobs, technology risks, programs undertaking ethical decision-making	Applied approach; three points as guidelines are offered for developers, primarily to avoid harm for humans	Partially (implicitly)	Yes (developers)	No	Yes: politicians and governments through laws and regulation to limit unwanted changes

Table 1 (continued)

Studies	Concept	Focus	Approach/framework	Actors with accountability			
				User	Manufacturer	AI robot	Other
Westerlund (2020)	An ethical framework for AI robots is offered based on previous scholarly roboethics literature	Design and use of AI robots but also "machine ethics" that indicates ethics relating to forms and codes of conduct for the AI robots	A framework of ethical perspectives to smart robots is offered in reflection to "Robots as object of moral judgment" and "Ethical agency of humans using smart robots"	Yes (both at individual and society level)	Yes	Mostly not, but leaves space for the potential of AI robots to become as moral and active agents as a potential scenario	n/a
Wright and Schultz (2018)	Social contracts and AI robots	Who the stakeholders are, what their expectations are, how they are impacted, and how the violation of stakeholder expectations can be avoided	Social contract theory. The study offers an overview of the factors influencing the likelihood and immediacy of automation-based labor disruptions and offers an integrated ethical framework with focus on the development of social contracts	Yes, implicitly as stakeholders As workers and consumers plus organizational users, but they appear more as receivers/impacted actors	n/a	No	Yes, nations and governments
Moor (2006)	Machine ethics	The existence and role of machine ethics	A variety of approaches is applicable to machine ethics as well as deontic/epistemic/action logics	Yes, adults appear as full ethical agents	Software engineers	No	Yes, for instance, banks, in relation to intelligent financial systems
This study	The study uses descriptive ethical theory to elucidate how introducing AI Robots into specific contexts shapes how we understand implications for moral agency, locus of responsibility, and moral intensity	Moral intensity, locus of moral accountability, and accountability dispersal. The use of AI robots can be illegal, immoral, permissible, and supererogatory	Normative ethics, business ethics. Different concentrations of accountability are outlined without the provision of new ethical norms	Yes	Yes	Not yet	Yes, governments and policymakers

and economics towards a comprehensive theory of rational action and social choice theory (Arrow, 1973). The *market failures approach to business ethics* (Heath, 2014) seeks to formulate normative standards implicit in the basic economic assumptions underlying the market economy's institutional mechanisms. It states that business and innovation require different rules than ordinary morality. The market failures approach sees market competition as driving the efficient allocation of goods and services to achieve greater common good (Heath, 2011). It understands regulatory and ethical intervention as levers to correct imperfections and thereby reduce any misallocation of resources (Heath, 2014).

Within normative business ethics, we draw on *the market failures approach to business ethics* in distinguishing between illegal, immoral, and permissible use (Heath, 2014). We complement these pillars with another normative business ethical category, which is the group of supererogatory actions (Driver, 1992). The rationale for using normative business ethics is its connection between practice and ethically ideal scenarios. In respect to the recent developments in AI robots, disregarding practical considerations would hinder the ethical evaluation of their use. *Supererogatory* actions represent an extra mile from what one expects morally (Driver, 1992; Mazutis, 2014). The three ethical categories outlined in the market failures approach are as follows: (1) *Illegal* is any action that is against the law and regulations; (2) *Immoral* is any action that only reaches the legal threshold's bare minimum. Statements such as 'we didn't break the law' signal illegality and suggest discomfort with the morality system (Wilson & Series, 2002). Bardy et al. (2012) define morality as the set of prevailing behavior standards that facilitate cooperative behavior; and (3) morally *permissible* actions are those not requiring explanations of putative fairness or appropriateness.

In normative business ethics, and more specifically in the market failures approach to business ethics, the subject (object-specific actions) can act upon the act or without a subject. Heath (2014) discussed object-specific unethical actions such as seeking privileges by using corporate assets or gaining insider information as a result of one's strategic position. Ethically debateable non-object-specific actions include, for example, exerting abusive behaviors. To achieve a relative balance between suggestions on what 'should' and 'should not' happen, we included the supererogatory category that is meant to encourage rather than discourage certain actions. Together with this fourth group, our proposed new framework will consist of two prohibitive (illegal and immoral) and two non-prohibitive (permissible and supererogatory) categories. As captured in Fig. 1 and the discussion, we illustrate further below how one can interpret these categories in relation to the four accountability clusters.

Extant ethical discussions typically draw on the binary logic that distinguishes between ethical and unethical acts and decisions. Bommer et al. (1987) identified different factors such as corporate goals, the juridical system, and religious or societal values that can encourage individual

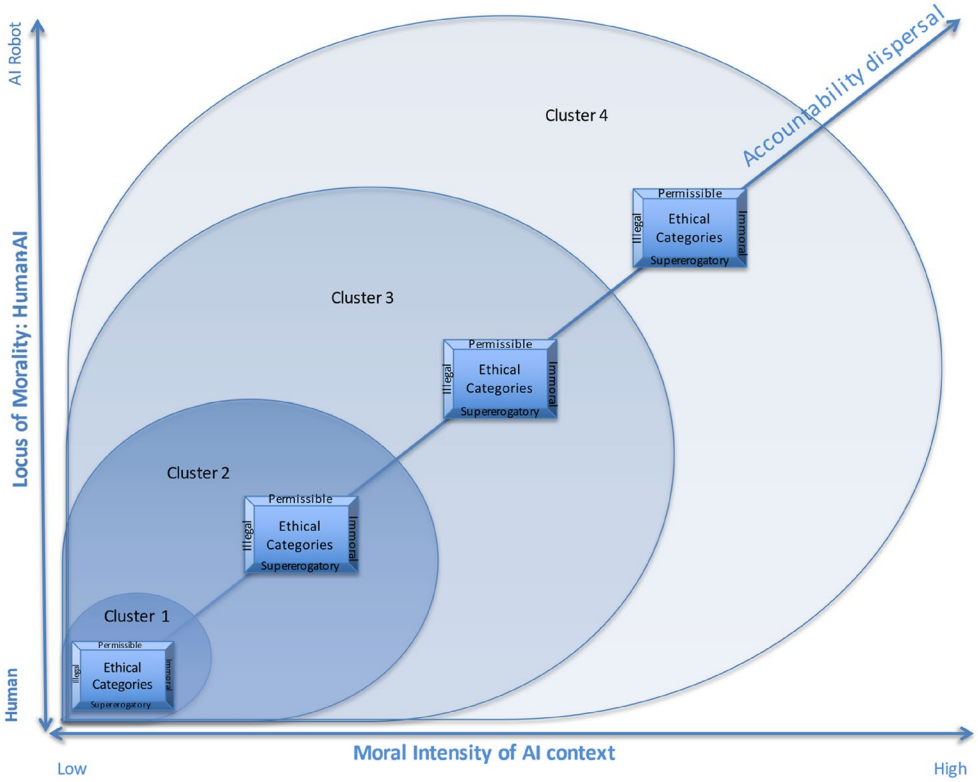

Fig. 1 New framework for AI robot accountability

decision-makers towards ethical and unethical decisions. Ultimately, however, researchers classify decisions into the binary groups of ethical and unethical. Similarly, Constantinescu and Kaptein (2015) explore various drivers of behaviors and question the extent to which individuals or organizations can be made responsible. However, pertaining to the outcomes, these authors also stick to the binary categorization of ethical and unethical dealings. Khalil's (1993) research examines ethical decision-making in the context of expert AI systems. He builds his argument by assuming that the decision-maker can choose among several actions that require evaluation as right or wrong, ethical or unethical. Khalil identifies reasons for ethical concerns. For example, expert systems lack human intelligence, emotions, values, and possess certain bias. He presents a variety of drivers, yet classifies the outcomes either as ethical or unethical. These studies provide useful insights into the underlying mechanisms and factors that influence decision-making when it comes to individual and organizational contexts. Thus, they are useful for individuals, companies, and governments to review their decision-making processes. However, we argue that from an ethical viewpoint, besides the drivers of decision-making, the outcomes also are relevant. Instead of a rather simplistic binary ethical/unethical categorization, there is space for a more refined approach consisting of the illegal, immoral, permissible, and supererogatory ethical categories. Research can benefit from having more options than the two extremes for ethical evaluations, especially when it comes to AI robots, where technological advancements and the different use of technology are diverse to the extent that their categorization into ethical and unethical categories became increasingly challenging and misaligned with practice.

A New Framework for AI Robot Accountability

The new framework can inform the ethical evaluations and subsequent action planning of managers, public policy makers, and civil society groups to better understand the implications of, and accountability responses to, AI robot applications. This is not intended as a prescriptive or static framework, given the inherent variation in application forms, movements in the state of technology and, crucially, shifts in societal expectations of what is and is not morally acceptable and legitimate (Suchman, 1995). We suggest two axial themes driving the framework—locus of morality and moral intensity—that combine in unique ways to render specific 'clusters of accountability' necessary for AI applications in business (Fig. 1). Figure 1 captures the increasingly dispersed nature of accountability, as an outcome of rising moral intensity and AI agency, and how this provokes different kinds of actors and levels of analysis. Our four accountability clusters correspond with the types of actors present at a particular level. Thus, in situations of concentrated accountability (i.e., low dispersal, low moral intensity, and human agency) AI robot accountability may be affected through well-defined and local actors. These could include the supplying AI designer company and the implementing company and user/s. An example of this might be deploying AI cleaning or even stocking robots within large warehouses. Contrastingly, in situations of widely dispersed accountability (i.e., high dispersal, high moral intensity, and low human agency), accountability clusters may draw in numerous, formal and informal actors across macro-institutional and supra-territorial arenas in order to provide accountability. An example of these kinds of AI robot settings could be international peacekeeping, military and/or humanitarian applications where deployments require accountability across different geo-political and legal arenas.

The Accountability Challenge of AI Robots

Figure 1 highlights our connection between the locus of morality and moral intensity in a context-specific AI robot application setting, along with the corresponding accountability mechanisms likely required (which will be discussed in more details further below). Drawing this link in relation to AI robot applications is necessary for the following reasons: to start, in a non-AI environment, if a manager or employee were to make an unethical decision or engage in an illicit practice (e.g., harassment, discrimination, deception, theft), the question of whom should be accountable is likely fairly concentrated (e.g., local/proximity to the person, department, or organization) such that the individual decision-maker—the moral agent—may be held solely responsible. The administration of any punitive or restorative accountability mechanisms also is likely to be local (line manager, human resource management, training programs, whistle-blowing procedures). Crucially, the focus and scope of these accountability mechanisms are not widely dispersed. However, introducing AI robots into such settings greatly complicates the interrelationship between the would-be wrong-doer (or unconsciously/complicit 'wrong-doer' for that matter) and corresponding restorative mechanisms, which the technological opacity associated with using AI robots exacerbates.

In the event of an ethical issue (perhaps in error) that causes harm directly, or facilitates harm indirectly, the question of accountability is far more complex. Is the AI robot, supporting AI system, developer, implementing organization, overseeing manager, industry regulator, or government responsible? Martin (2019, p. 129) broached this problem with regard to the question of accountability for the un/intended consequences of algorithms, whose

agency varies from *"simple if–then statements to artificial intelligence (AI), machine learning and neural networks." Because of this, Martin finds (2019:130) it inevitable that "all algorithmic decisions will produce mistakes" that, if left undetected, could reproduce unfairness, inequality, and harm for different stakeholders. Thus, even if 'Roboethics' are designed-in, for example, to inhibit AI robots' unethical decisions, ethical problems may emerge and persist that disperse accountability throughout a range of actors and activities beyond the initial designer. We thus anticipate something akin to 'Algorithmic Accountability' to occur in the context of AI robot applications and urge discussing the context-specific mechanisms for accountability that would include, but extend well beyond, the initial coding role of the AI designer or user (*Table 1 *suggests a preoccupation with designer-user accountability). Therefore, the next sections discuss key drivers of accountability of AI robots' use for the business ethics and wider management studies community*. Table 2 connects the 'accountability clusters' with the ethical categories of illegal, immoral, permissible, and supererogatory. Each intersection is illustrated with examples of *AI robots' use for different purposes and in different settings.*

Locus of Morality

AI robots' use influences the extent to which the locus of morality, defined as the autonomy to choose an ethical course of action (Bruder, 2020), is mostly concentrated in human agents (weak AI agency) versus human agency being less straightforward with concentration towards AI robots (strong AI agency). Strong AI agency does not imply the lack of human agency but instead the hidden nature of human agency. Recent research from the field of 'Roboethics' (Leventi et al., 2017) indicates that AI robots may not only share agency with humans in application settings but may learn from them and 'improve' (machine learning). Advancements in robotics have led to the emergence of 'smart robots,' which Westerlund (2020, p. 35) defines as "autonomous artificial intelligence (AI) systems that can collaborate with humans. They are capable of "learning" from their operating environment (…) in order to improve their performance."

Theoretically, repeating poor human decisions, not noticing certain harms or injustices or even unwittingly causing them, are all possibilities of escalating AI moral agency. In building our vertical axis (Fig. 1) we combine the notion of the locus of morality with Martin's (2019, p. 131) assertion that "*Algorithms relieve individuals from the burden of certain tasks, similar to how robots edge out workers in an assembly line. Similarly, algorithms are designed for a specific point on the augmented- automated continuum of decision-making."*

Table 2 Accountability clusters and ethical categories

Actors Ethical categories	Cluster 1 Professional norms	Cluster 2 Business responsibility	Cluster 3 Inter-institutional normativity	Cluster 4 Supra-territorial regulations
Illegal (Heath, 2014)	Privacy breach through disorderly design	Stealing data from competitors with the help of AI robots	Algorithmic injustice in immigration systems	The development and use of LAWS
Immoral (Heath, 2014)	Design that deceives people about communicating with a human rather than an AI robot	Artificial shortening of product lifecycle through production by AI robot	The pervasive use of social auditing systems for political purposes	Use of killer drones, even when the use of such drones is legally permitted
Permissible (Heath, 2014)	Less human touch if overall quality of elderly care can be enhanced	Job loss due to increased use of AI robots	The use of AI tacking systems in a pandemic	Using AI robots for self-defense purposes
Supererogatory (Driver, 1992; Mazutis, 2014)	Improved analytical accuracy in medical diagnosis with the help of healthcare AI robots	Increased health & safety through using AI robots for dangerous tasks	AI in smart cities in support of sustainable growth	Using AI space robots to detect flood and save lives and wealth

We note here that the more extreme ends of the upper continuum—AI robots as autonomous ethical decision-makers—are at present theoretical. Full autonomy is a hypothetical scenario that serves as an orientation point rather than an actual, attainable quality of AI robots. This is subject to the state of technology and social license around AI acceptance over time, and even strong AI ethical agency may well experience some 'bounded' autonomy. However, Westerlund (2020) informs us about the potential for a broad spectrum of AI agency, ranging from robots as passive recipients of human ethics (the object of programmed ethical codes) to highly active agents (subjects of ethical judgment). Westerlund (2020) also suggests that AI robots may become recipients and shapers of the socio-cultural ethical norms at a more macro-social level over time. Thus, the locus of morality is likely to shape both 'local' accountability responses, emphasizing the role of designer-user accountability solutions (Bench-Capon, 2020) as well as broader macro-social transformations that may require temporally undefined responses by a constellation of organizational, governmental, and civic agents and structures.

These considerations have informed how we structure the vertical axis in Fig. 1—the 'locus of morality.' In normative ethical theory, a primary assumption is that ethical decisions about the respective harm or freedom resulting from an action (right rule/best outcome) is conditional upon a rational choosing agent. Depending on the strand of moral philosophy, this could involve necessary human characteristics such as capacities for moral reasoning about rights and consequences of actions, a socially acquired sense of virtue and moral character as well as, more existentially, some innate notion of a moral impulse, empathy, and/or care for others. The latest research in business ethics suggests that managers can call upon their 'moral imagination' (Johnson, 1994), including both reasoning, empathy, and sheer intuition, in reaching the best possible ethical decision (Tsoukas, 2020). The current force of technological development, at least for some AI robot applications, is towards human imitation, including, especially, the way we make choices. This would most certainly include choices of an ethical nature (or outcome), ranging from compliance and imitation to, potentially, autonomous judgment. There is a moral risk of imitation, however, while "AI system is a tool that will try to achieve whatever task it has been ordered to do. The problem comes when those orders are given by humans who are rarely as precise as they should be" (Kaplan & Haenlein, 2020, p. 46).

Thus, the vertical axis captures on a continuum from human to AI robot, where the *locus* of moral decision-making increasingly resides. At the lower end, what others have called 'Assisting AI,' such as AI-assisted diagnostics, humans largely set ethical parameters for AI robots, who in all cases would make the final judgments regarding ethical decisions. Humans can program ethical codes and make any necessary adjustments. The AI robots cannot do this. The boundaries between human and AI robots as the origin (locus) of ethical decisions become increasingly blurred as we move up the continuum, where both humans and AI robots may assume different amounts of autonomy over ethical decisions. The top of the continuum represents a theoretically possible position (Westerlund, 2020), where humans have almost no part in any ethical decision-making, leaving this entirely to the AI robot and questioning, ultimately, who is accountable for an un/ethical decision executed (or not) in this context (e.g., the system, the product, the company or the government).

Moral Intensity

This leads us to the second determinant of accountability, that of *moral intensity*. Moral intensity of a given situation has been well documented in the descriptive ethical theory literature, most notably with Jones' (1991) issue-contingent model showing how perceptions of moral intensity affect an individual's decision-making. As we are concerned with how accountability clusters develop for particular AI robot application settings, we deploy moral intensity in a different way from Jones, defining it as the context-specific exigencies of *vulnerability* and *scale* that amplify un/intended consequences in AI robot application settings (i.e., the focus here is not the decision-making subject's perception of moral intensity). As we explain later in our four clusters of accountability, moral intensity may stem from (any combination of) numbers of humans potentially effected, the vulnerability of human agents and/or the severity of current and legacy effects on a community or ecosystem. As our approach is a descriptive one, we will not discuss here how designers from deontological or teleological frameworks formulate AI robots' decisions (see Bench-Capon 2020 for an explicit discussion on these).

In ethical theory, including normative ethics, researchers commonly emphasize the moral intensity of a decision. For example, the magnitude and distribution of consequences—both beneficial and harmful—is the task of utilitarian ethical theory. Theories of justice recognize the relative vulnerability of certain actors over others, rendering them more vulnerable to receive potential harm or have access to benefits denied. In each case, harm/benefit has a certain magnitude or intensity to the decision-making. In Heath's conceptualization, moral intensity comes with increased interaction: "iteration of the interaction only intensifies their [individuals'] incentive to act in the same [morally questionable] way" in competitive settings (Heath, 2007, p. 360).

Overall, there is a spectrum of situations between low and high morally intensity (Jones, 1991). Factors such as time, magnitude, proximity, and distribution of consequences can

play a role. An individual employee regularly taking longer rest breaks than his/her colleagues is on a different moral intensity compared to a hospital that prioritizes profits over patient safety in the long run. Using examples in our framework, we can similarly argue that a faulty AI cleaning robot, for the most part, will result in comparatively benign outcomes (getting briefly lost) than will an AI robot that fails in the correct diagnosis in health service encounters.

Accountability Clusters for AI Robots

While we have already begun to discuss AI accountability above, first it is necessary to draw a detailed link between our theoretical constructs and their corresponding accountability clusters as Fig. 1 presents. The locus and intensity of morality in AI robot settings necessitate special consideration of accountability and governance. Although concerned principally with algorithms, Martin (2019) underlines the inevitability of decision mistakes (both human and system). Intentionally or otherwise, occasionally managers will make poor decisions. The issue here, and of pressing concern for business ethics scholars, is what happens when mistakes do occur (e.g., mis-reporting a company's financial health): "*Ungoverned decisions, where mistakes are unaddressed, nurtured, or even exacerbated, are unethical. Ungoverned decisions show a certain casual disregard as to the (perhaps) unintended harms of the decisions; for important decisions, this could mean issues of unfairness or diminished rights" (*Martin, 2019*, p. 132).*

We argue that while AI robots' capability to collect and record information that may indicate mistakes (or causes of) is extremely powerful, the capacity for identification, interpretation, judgment, and deliberation may be correspondingly minimal. Weak AI ethical agents may not recognize unethical decisions that require correcting. Moderate AI ethical agents may imitate and repeat them (as 'good'). Strong AI ethical agents may overlook serious harms (e.g., via lack of empathy) in pursuing other 'good' organizational ends. In short, AI robots' decisions and practices that precipitate harm, inequality, and unfairness may go uncorrected over time. Thus, it is necessary for stakeholders to think seriously about accountability issues in specific AI robot settings. In response, we suggest four accountability clusters as indication of actors, resources, and activities to address accountability. In short, accountability requirements, as determined by the locus of morality and moral intensity, may be local, concentrated, and ad hoc (organizational supervisor), or widely dispersed across private, public, and civil society agents in an ongoing discourse of accountability (Buhmann et al., 2019). There will be markedly different requirements for accountability clusters between, for example, situations where the locus of morality is mostly concentrated within human agents (weak AI ethical agency) and where moral intensity is low (e.g., office cleaning) compared with situations of stronger AI ethical agency and where moral intensity is far higher (e.g., AI carers or soldiers). However, we do not provide new ethical norms (e.g., principles for managerial accountability per se) but indicate context-specific accountability clusters that may well include norm-making administrative mechanisms.

In this section, we respond to Buhmann et al.'s (2019) warning from algorithmic research that there may be special, discrete accountability characteristics specific to AI and system learning technologies (such as technical opacity) that render expectations of accountability as highly fluid and nuanced. Rather than provide a rigid typology of accountability mechanisms that managers or policy makers must follow, we interpretively develop clusters that fall loosely into the four different clusters of accountability in Fig. 1. In a sense, we demonstrate here how managers or policy makers might use our theorization in attempts to combine scenario planning around AI robot accountability. From our theorization, we developed four clusters (see Fig. 1) that delineate context-specific applications of AI robots. This enabled corresponding consideration of appropriate ethical categories. Note here that the depicted clusters are not mutually exclusive but cumulative, each of them is nested inside the other, like Russian (matryoshka) dolls, incorporating an increasing number of agents as moral intensity and agency rises. For each accountability cluster, we will also discuss the normative ethical properties of illegal, immoral, permissible, and supererogatory. We characterize the ethical dynamics and corresponding accountability clusters, providing further corresponding examples of AI applications (see sources in relation to the examples in Table 3).

Cluster 1 Professional Norms

Cluster 1 represents a relatively local and concentrated accountability cluster, characterized by applications with low AI ethical agency and low moral intensity where questions of accountability are largely contained to a well-defined designer–device relationship. Akin to safety certification schemes for products, designers make ethical decisions and then encode them into non-reflexive task and behaviors that AI robots can and cannot do. We distinguish here between AI robots as imitators (such as certain chat bots that support booking processes), and AI robots in this context that follow simply pre-programmed codes of conduct (such as smart heating systems). Considering that designers shape the technical features and may incorporate ethical considerations, professional norms play a key role here, especially considering that designers are unlikely to receive much pressure from governments and international bodies on AI robots' use. It is at this lowest level of agency and intensity that we would situate models mainly pertaining to *supererogation*.

Table 3 Illustrative cases of AI robot applications in different contexts

AI robot example	Illustrative case/source	Locus of Morality	Moral Intensity	Accountability dispersal
AI cleaning and gardening robot	Court case number of a rare case of a permanent eye damage due to lost control over AI-enhanced lawn-mover: Nanmuzzi v. King et al., 660 F. Supp. 1445, United States	Clearly identifiable human moral agency. Primarily the designer due to a design issue	The moral intensity of the use of AI cleaning and gardening robots is typically low	Accountability dispersal is typically low and depending on the case, the relevant persons are mostly identifiable
Smart heating systems and smart energy use	Smart home solutions are encouraged by the UK Government as part of the "Buildings Mission" (Heat in Buildings; www.gov.uk)	Clearly identifiable human moral agency (level of heating is set up by human user)	Low moral intensity	Accountability dispersal is typically low
Explorer AI robot (e.g., mining, deep sea, space)	The Convention on International Civil Aviation, US, regulates coordinated access to outer space, relevant also for AI robot applications	Human moral agency is relatively easily identifiable	Moderate moral intensity, except for e.g., military applications	Accountability dispersal is moderate
Agricultural AI robot	AI robot device for Pyralidae insects—IntechOpen (see reference Hu et al., 2018)	Relatively easily identifiable human moral agency. Typically organizational users	Rather low moral intensity but appropriate supervision is still required	Accountability dispersal relatively low and can be narrowed down to a limited number of organizational actors and professionals within the affected organizations
Repair AI robot	§ 87(1)(6) BetrVG (Germany) regulates the use of technical surveillance devices is smart factory (inc. repair) settings	Relatively easily identifiable human moral agency. Typically organizational users	Rather low moral intensity but appropriate supervision is still required, e.g., to avoid industrial spying	Accountability dispersal relatively low and can be narrowed down to a limited number of organizational actors and professionals within the affected organizations
AI-supported data management (e.g., healthcare and social care, criminal records)	Medical Service Act, Article 53 (Evaluation of New Medical Technology) in South Korea regulates relevant information sharing	The identification of human agency can be relatively difficult in certain cases	High moral intensity; high risk harm	Accountability is dispersed across various organizations
Driverless cars	Insurance legislation based on accidents: Hilgendorf: Automasiertes Fahren und das Recht, Deutscher Verkehrsgerichtstag, 2015, Vol. 53, p. 66	It can be very difficult to identify human agency. However, sever intoxication (§316 of German Criminal Code incurs criminal liability of the human behind the wheel)	High moral intensity; high risk of physical harm	Widely dispersed accountability that can include user, designer, manufacturer, and governmental and other regulatory bodies
AI robots in elderly care, in rehabilitation	Stevie the Senior Care Robot, Time, "The Robot That Could Change the Senior Care Industry," Time (time.com, 14.10.2019)	It may be difficult to identify human agency (especially by the user). Regulations are yet to address to the extent to which robotic assistance can replace human interaction	High moral intensity, high potential of emotional attachment on the user side towards the AI robot that can further increase human vulnerability	Widely dispersed accountability that can include user, designer, manufacturer, and governmental and other regulatory bodies

Table 3 (continued)

AI robot example	Illustrative case/source	Locus of Morality	Moral Intensity	Accountability dispersal
LAWS and other military AI robots	"Autonome Waffen: Wie Roboter den Krieg menschlicher machen sollen" by Sharkey in Spiegel, 26.06.2014. It is argued that military AI robots are unable to distinguish between whether a human would survive or die in certain settings and allocating this decision to AI robots is inhuman	It could be very difficult to identify human agency if military AI robots were used for fight. AI robots' use in self-defence situations is considered in Section 32, §3 of the German Criminal Code	Extremely high moral intensity; high risk of physical harm	Widely dispersed accountability that can include user/soldier, designer, manufacturer, and governmental and other regulatory bodies

For instance, smart heating systems achieving environmentally friendly solutions without imposing significant risk on stakeholders is a good example. However, 'low' risk does not imply 'no' risk of ethical issues arising, implying that a significant degree of accountability is always required, albeit more locally administered in these settings.

Low moral intensity and a locus of morality that is closely attributable to humans characterizes this cluster. We include smart heating and cleaning systems at homes in this category: humans set the exact activity details (for instance, the cleaning route and sequence) and we consider the activity type typically harmless. Immoral or illegal use of these AI robots appears to be unlikely, thus their moral intensity is limited. The locus of responsibility is primarily with humans (e.g., to select the degree to which the property would be heated). Overall, this group represents low ethical risks, which where appropriate, require human supervision, even if these measures play more of a preventive role rather than managing previously experienced issues with certain AI robots. It is difficult to construe illegal applications of, for instance, smart heating and cleaning systems, but we cannot exclude it as a possibility that some smart systems have a reprogramming capacity for the unsolicited monitoring of someone's private life or business matters besides their original purposes (e.g., cleaning, heating). A potential immoral application is when someone misleads others about his/her inability to attend an event due to compromised mobility and attends another event instead, with the support of a mobility AI robot. An example for permissibility in this category is that a cleaner may need to seek other work if his/her client sets up a smart cleaning system at home or workplace.

Cluster 2 Business Responsibility

Cluster 2 represents an accountability cluster characterized by moderate moral agency (with the locus of responsibility still closer linked to humans/groups of humans) within contexts where moral intensity is moderate. This could mean contexts where there are few humans or, for instance, the nature of the task poses little threat to humans or ecosystems, despite an increased level of AI agency. Interestingly, application contexts that might prove difficult or impossible for complex organic life to operate in, such as mining, deep sea, or space exploration, could invite AI robots with considerable degree of autonomy. The relative lack of ecological or human threat would likely result in a more concentrated cluster of mostly professional and/or organizational-level actors within temporally bounded moments (e.g., user organization following pre-existing industry regulations). This makes 'business responsibility' characteristic in this cluster, which refers to the liability of the organization that uses the AI robot. Moreover, it is in this cluster that we might see potential *permissible* decisions, practices, and

outcomes. An emphasis on setting clear parameters for AI robots based upon organizational values, goals, mission, and codes of conduct would most likely complement designer-lead AI robot ethics.

Agricultural AI robots for insect detection are part of this group with nearly full autonomous operations that have low physical or other risks to humans. In addition, AI robots intended for weeding and seed-planting belong to this category. Similarly, repair and inspection AI robots can enter spaces that humans would struggle to reach and use sensors that accelerate the sensing capacities humans have. For example, the flexible elastomeric two-head worm AI robot imitates inch-worm motion, holds sensors that explore their environment and learn repair-work patterns. Companies can use it for repair and inspection (Tahir et al., 2018). The permissibility of such AI robots lies in that although larger groups of workers may lose their jobs, the humans who continue the work in AI robot-enhanced environments can enjoy improved work conditions (Fleming, 2019). With fewer workers in an AI-robotized environment, there is a reduced risk of workers engaging in dangerous tasks. The increased safety element represents supererogation. Unlikely potential illegal applications include causing harm to humans due to negligence or as a planned criminal act, or the even more efficient harvesting of illegal drugs with the help of AI robots. An immoral application may be to pressurize the human workforce to 'compete' with the AI robots' performance carried out in different sites of an organization. The presence of a 'supervising' human agent may still be required to provide ad hoc and/or strategic monitoring to ensure alignment with industry codes, as well as correcting any ethical 'blind spots' that designers or organizations may have.

Cluster 3 Inter-institutional Normativity

Cluster 3 reflects situations of relatively high AI ethical agency coupled with relatively high moral intensity. Accountability may be relatively dispersed between actors subjected to institutional norms of, for example, a regional industry and/or national context that prompt the need for interorganizational liaising on ethical implications. Regulatory, industry, trade union, and civil society institutional actors, for example, might be present but in a national or region-specific context. Inter-institutional normativity refers to the nature of decision-making processes in which one concludes actions and outcomes to be ethically desirable. In this cluster the interaction between different organizations plays a significant role (instead of the focus on a single organizational setting). AI robot applications in this group deserve special attention to minimize occurrences that involve *immoral* decisions, practices, and outcomes. While we do not exclude legal mechanisms altogether (after all laws have a moral basis), we emphasize here the focus upon institutions—including certain governmental bodies—relevant for how industry uses AI robotics. We might anticipate the content of institutional norms to vary according to the geo-political contexts; however, the presence of institutional norms would likely reflect some kind of social contract to protect citizens in contexts of heightened vulnerability.

Examples in this cluster are the use of AI-supported healthcare data management systems (with the need for interorganizational liaising between professional healthcare bodies, programmers, and the government) as well as AI-supported crime-prevention systems (typically operated at the national level, even though international collaborations are increasingly important for crime prevention). Humans may maintain a strategic control of care planning and resource allocation decisions in the healthcare and crime-prevention data management systems, and then they implement these decisions in AI robots' daily encoded tasks. Similarly, AI robots applied for social auditing that can encourage social distancing and other relevant safety measures belong to this group. Potential ethical transgressions are less likely to originate from the AI robot itself (as it is an imitator), but the lack of AI robot judgment means that any un/intended consequences of poor human decisions may go routinely unnoticed and perpetuated by AI robots. The lack of AI reflexivity could perpetuate unfairness, inequality, social exclusion or even harm, in various learning and rehabilitation environments. Strong normative institutional accountability mechanisms need to be in place to not only set the parameters for actors implementing AI robots in such settings but to measure and provide feedback upon shortcomings against a set of agreed norms. An example for when accountability mechanisms may not have been in place is when UK-wide databases of more than 400 thousand criminal records, including criminals' fingerprint information, have been deleted (Reuters, 2020). Deleting criminal records could have occurred in the case of non-AI-enhanced systems as well, but AI robots can accelerate the speed and volume of this data-specific damage that is vital from a social security perspective. The absence of appropriate data-monitoring mechanisms raises the question of immorality at institutional levels (in this case the police). We cannot exclude entirely the possibility of illegality. For example, in a hypothetical scenario, a police officer who had access to the compromised criminal records system might want to cover up someone's criminal activities, but this is yet to be consolidated as the investigation is under way.

Cluster 4 Supra-territorial Regulations

Cluster 4 represents a cluster of applications, which we describe by strong AI ethical agency, high moral intensity, and the widest dispersal of accountability between actors.

As a result, accountability clusters are likely to be fluid and complex, requiring ongoing discourse between designers, users, organizations, industry, regulatory, and thus, supra-territorial regulations. Multiple actors may compromise clusters of multiple actors with a range of alternate vested interests (e.g., national and regional government, national and international law, civil society, industry bodies and corporations). Supra-territorial regulations refer to the need for collaboration between individual and organizational actors at an international level. It is at this layer that we might have situations that result in *illegal* decisions, practices, and outcomes. An emphasis on strengthening and overseeing regulatory mechanisms at the highest level (e.g., including international legal apparatus, media, and civil society) might be necessary to complement more local and regional mechanisms. Examples of AI robot applications falling into this category could involve certain health and care services, and military application contexts, especially where there are a high number of affected people and/or the nature of the application implies significant vulnerability.

The high level of accountability dispersal does not imply that the AI robots 'usurp' the role of ethical human decision-making but it is becomes increasingly difficult to attribute AI robots' acts to specific individuals or organizations. If not managed properly, illegal use of AI robots is likely to occur in this group. AI robots in this group can be increasingly autonomous. An example for very high moral intensity, where also the perceived locus of morality falls far from individual human beings, is using Lethal Autonomous Weapon Systems (LAWS). The operational autonomy is very high with very little to no human involvement. The agreement about avoiding or allowing certain use of these AI robots is international in nature as the nature of defense typically is (apart from internal conflicts). Besides LAWS there are other considerably autonomous AI robots such as military drones and Big Dog (Lin et al., 2014) that is considered as a carrier of military equipment instead of an attacker robot, yet still in support of war effort. Highly autonomous AI killer robots make decisions on their own—we could consider their manufacturers as facilitators but according to Byrne (2018), not as murderers themselves. Intergovernmental regimes are required to collaborate to hinder the illegal use of military AI robots. Depending on the regulatory settings, the use of LAWS is typically illegal. In the absence of legal prohibition, they may be immoral. Using other military AI robots may be permissible (e.g., for self-defense purposes) and even supererogatory (e.g., to save lives in a natural disaster).

Driverless cars are another example of autonomous AI robots, where the locus of accountability is not primarily with the human behind the wheel. The driverless car sets out with a program that incorporates speed limit guidance but learns that other cars exceed the limit and concludes that it should speed too. Tragically, there were different incidents where a Tesla car traveling over the speed limit resulted in deaths (Etzioni & Etzioni, 2017). This caused Tesla to further examine its autopilot driving system. The nature of the regulatory environment for driverless cars is increasingly international as they are becoming an inherent part of international mobility. While country-specificities relevant for driverless vehicles may apply (e.g., the lack of a speed limit on the German highway (the Autobahn), there is an increasing need for consistency at a supra-territorial level (similar to permitting using EU driving licences in any of the member states). Further, while using AI robots in care homes can increase elderly life quality (Broekens et al., 2009), it also implies some risks, especially where it is unclear who should 'supervise' these AI robots or when assigned supervisors neglect checking on the AI-enhanced care robots and the compliance of their use with international health and safety standards.

Dynamic Contextual Factors

There are some dynamic factors that we need to highlight as caveats to how to interpret Fig. 1. The dimensions we suggest will be subject to movements over time depending upon the actual context-specific application. The key trends that may influence applications of the framework include the *changing state of technology*, for example, with machines varying in terms of the extent of human imitation they possess (analysis-intuition). Another factor is the relative *labor/skill displacement* (Wright & Schultz, 2018), where certain (low skill) workplaces are potentially decimated depending on the level of imitable specialisms in the workforce. Reduced employment opportunities and AI-supported warfare among countries trigger equality-related concerns as well. "Unless policies narrow rather than widen the gap between rich, technologically-advanced countries and poorer, less-advanced nations, it is likely that technology will continue to contribute to rising inequality" (Wright & Schultz, 2018, p. 829). Finally, there is a factor of *unknown outcomes*. For example, currently we do not know whether AI robots will make better moral decisions than humans, or more consistent ones. We also do not know if they may be able to redefine the moral parameters independently. For instance, it appears that AI has the potential to both reinforce and reduce racism (Noble, 2018). AI robots can learn, for instance, swear words and bullying behaviors (Dormehl, 2018). While such behaviors may not be illegal, we can regard them as immoral and thus, policymakers should support developing appropriate monitoring mechanisms. It is noteworthy that although attention is diverted from the technical innovativeness of AI robots towards corners around their interaction with humans, AI robots do not develop immoral behaviors themselves but learn those from humans,

for instance, through pre-programming or the imitation of humans.

Discussion

Theoretical Implications

This study offers a new framework for AI robot accountability that conceptualizes AI robots' ethical implications along the dimensions of locus of morality and moral intensity. Considering that an AI robot may have a potentially high degree of decision-making discretion (much like a human employee—after all, imitation and learning from humans is among the goals of AI robot designers), in the event of an accidental error, misconception or even a well-intended misinterpretation of the data-response that causes harm directly or facilitates harm indirectly, the question of accountability may be widely dispersed (among for instance the developer, maintainer, the implementing organization, overseeing manager, informed by industry norms and regulations). We argue that this significant ethical deviation in accountability needs acknowledgment and exploration in a context-specific way. This study identifies accountability clusters that we can characterize by different concentrations of accountability—without the provision of new ethical norms (e.g., principles for managerial accountability)—that can inform norm-making administrative mechanisms.

The study draws on normative business ethics, especially the market failures approach to business ethics (Heath, 2014), when it comes to describing moral intensity. Heath's work revisits the unanswered questions of organizations' ethical responsibilities, and which considerations management should consider to ensure ethical operations. It is noteworthy that the market failures approach to business ethics has grown from the heated debate between shareholder and stakeholder theories (Young, 2015). Milton Friedman states that the responsibility of business is to meet shareholders' need by increasing profits (Friedman, 1970). On the contrary, stakeholder theory argues that the firm's goal is to act in the interests of all their stakeholders, not only in their shareholders' interests (Freeman, 1994). Heath (2014) claims that organizations should avoid distorting competition by focusing only on profit maximization. The ethical categories of illegal, immoral, and permissible use stem from Heath's conceptualization on organizational action.

However, Heath's market failures approach to business ethics distinguishes only between acts that need to be prohibited either by law (illegal) or by following moral standards (immoral) or can be allowed (permissible). It lacks suggestions on what organizations should encourage to exceed minimum ethical requirements. Thus, this study extends the conceptualization into the broader family of normative ethics and integrates the ethical category of supererogatory use (Driver, 1992; Mazutis, 2014) to be able to offer insights not only on restrictions but also encourage certain development. Finally, this study views Heath's primary focus on firms as a limitation of the market failures approach because the ethical status of any occurrence appears to be dependent exclusively on companies. Besides firms, individuals (Soares, 2003) and governments have ethical responsibility too, as it has been highlighted in debates on the ethical implications of environmental problems (Fahlquist, 2009). Thus, this study broadens the conceptualization from a corporation-focus in a way that encompasses the ethical standing of individuals and the military.

The proposed framework has the potential to trigger further academic discussions on moral accountability and moral intensity and advances knowledge through the systematic combination of the two phenomena for AI robots' use. Regarding moral intensity, we consider illegal, immoral, permissible, and supererogatory use that encourages a non-binary approach towards the ethicality of AI robots' use. The position of certain examples of AI robots' use across the outlined clusters is fluid over time. For instance, driverless cars would already fall under the umbrella of supra-territorial regulations but are still rather close to interinstitutional normativity where they would have been located a decade ago. There is a time perspective to how the clusters develop because with time the stakeholders' position may change in reflection to ethical, social, and environmental matters (Longoni & Cagliano, 2018). Further, the group of stakeholders relevant for certain AI robots' use may widen or shrink, including designers, individual/organizational users, governments as well as intergovernmental regimes. Finally, Vogel (1992) acknowledges that the harmonization of ethical standards across different groups, regions, and countries is very slow, especially compared to the pace of technological innovation.

Regulatory Considerations for the Ethical Use of AI Robots

The challenge of moral intensity and moral accountability in using AI robots, if the pervasive nature of AI continues, creates a second challenge of how regulators should act to move towards an ethically appropriate application of AI. Regulators could use the proposed framework to inform policymakers' discussions on morality and accountability in relation to AI robot use cases. For instance, based on relative moral intensity and the locus of morality, policymakers can situate certain AI robot uses across the outlined clusters—as part of scenario planning—to review accountability dispersal and its future potential development. In addition, the framework can help in stakeholder mapping that identifies stakeholder groups that should be included in relevant conversations

currently and in the future. AI robot design experts may also benefit from using such a framework to enhance their ethical considerations and explore ethical possibilities in specific contexts. As the position of an AI robot use is not fixed but can move in between clusters, regulatory considerations could reflect on this. For instance, imitation-driver chatbots used even only in booking settings may "qualify" for a morally more intense cluster if the chatbot learns and applies bullying. The AI-enhanced data management issues of criminal records, for instance, may well be situated in the cluster described by interinstitutional normativity. It can evolve into the supra-territorial regulations with time, especially with the increasing need for sharing crime-prevention practices at an international level. The move between clusters of accountability requires including different stakeholders that the proposed framework can be helpful to support developing an in-depth understanding and relevant reflections.

We offer some consideration points for regulators and other decision-makers for ethically using AI robots regarding accountability and in reflection to moral intensity. In doing so, one should pay special attention to learning mechanisms and decision-making processes as both are vital for developing and using AI robots (Baskerville et al., 2020; Benlian et al., 2019). Regulators should pay special attention to not only what AI robots should learn but also to what they should avoid learning. In the case of driverless cars, for instance, learning algorithms should include restrictions to learning dangerous learning behaviors. While decision-making is almost fully automated, humans should have the option to revert to a less autonomous mode but in a transparent way, i.e., it should be tracked when a human is only a passenger in the driverless car and when they act as drivers, so that accountability remains clear, and humans cannot blame their wrongdoing on AI if they cause an accident. There are certain AI robots in this category that should be prohibited from use, for instance, LAWS. The implications of using LAWS impose high risk on people's lives, including civilians. Developing these autonomous weapons is beyond international peace agreements, yet some countries have invested in their development. A major ethical concern is that humans barely have any chance against highly accurate and intelligent killer robots (Byrne, 2018). Applying originally military AI solutions in non-military settings, for instance, for lifesaving in emergency situations (e.g., identifying and saving people and animals in the event of a major flood) should be further encouraged as these fall under supererogatory applications.

We suggest planning with a regular audit of the applied AI learning mechanisms because they may require updating with the innovation of medical procedures. The reliance on historical data has its limitations in formulating optimized solutions and we should acknowledge this at a regulatory level. For decision-making, approval seeking mechanisms from responsible contact persons should be arranged. Relying on AI-suggested healthcare and social care solutions without expert approval holds high ethical risks. However, AI robots can utilize historical patient data to inform healthcare and social care decisions and are exempt from biases. Note that risk assessment is important as it allows organizations (vis-á-vis hospitals) to find the acceptable balance between safety and avoiding dehumanization. Further, given heightened stakeholder precariousness and the desire for managers to fill short-term labor gaps, government level controls are necessary to prevent organizations from misapplying AI. This would ensure a portfolio/needs-based approach where AI capabilities match with human needs.

Learning mechanisms should include some restrictive features, such as processes that can harm human safety (including work practices and food safety), which one should avoid and, in some cases, even unlearn. For decision-making, management should monitor and review periodically the AI robot's role in decision-making to further improve situational response. Comparing outcomes with and before/without using AI robots is advisable unless the comparison compromises safety. For instance, one should still check regularly AI robots working in insect identification and elimination to ensure that they meet human safety regulations, even if we typically consider these AI robots as benevolent towards humans.

One should monitor and review regularly the AI robot's learning mechanisms to avoid potential ethical concerns, e.g., pertaining to privacy and data management. These issues are not entirely new but can be strengthened through using AI. Likewise, management should periodically review the level of AI decision-making, even if humans in this group exercise control. A challenge is to ringfence time for these evaluations when there is low ethical risk overall. However, awareness helps because low likelihood can still incur high-impact events, for instance, privacy and data management practices. For example, in a hypothetical scenario, a cleaning AI robot can collect confidential data from the building in which it is used, thus raising organizational/national security concerns.

Limitations and Directions for Future Research

This study follows Western ethical standards and consequently alternative explanations may apply internationally or even across different Western countries. Similarly, the variability of legal frameworks across countries may vary in an international regulatory context. The examples we used in this study derive from regulatory debates, court appeals and from previous papers and not from primary data collection. While we see this as a potential limitation,

Table 4 Future research agenda for the ethical use of AI robots

Future research topics	Future research questions
Illegal use of AI robots	How do legal systems across different cultural settings reflect on recent AI developments? Which approaches take a more reactive/proactive legal position and where do they put the emphasis? Where is the locus of accountability in current AI regulations?
Immoral use of AI robots	What is considered morally acceptable and immoral regarding using AI robots? How does AI change more traditional views on morality? How do AI robots influence social biases (e.g., in relation to skin color and gender) and vice versa?
Permissible use of AI robots	How does permissibility change over time and with the development of AI robots? How does the increased use of AI effect interaction between health, social care staff, and patients? To what extent is the loss of 'human touch' permissible and under what conditions?
Supererogatory use of AI robots	Why is supererogation overlooked in ethical evaluations of AI robots? How could supererogation be incentivized? What are the risks of supererogation (i.e., how wrongdoers of AI development/use may buy goodwill through additional supererogatory features)?
Accountability dispersal	Which consultancy and decision-making mechanisms should apply between governments at an international level as well as AI robots' designers, manufacturers, users, and other stakeholders to navigate in AI robots' use cases where high accountability dispersal applies? How could these mechanisms be codified as part of codes of ethics relevant for AI robot applications?
Locus of moral responsibility	How does hybrid/shared moral responsibility work between AI and humans? What is the level of moral responsibility that researchers can assign to AI? Which factors influence the location of the locus of moral responsibility? How can AI robots handle the choice between two ethically debatable options (e.g., whether an autonomous car should risk the lives of its passengers or the pedestrians in an unexpected situation)?
The role of government in developing AI robots	How do governments utilize the rise AI robots in both military and non-military settings (such as public services)? What is the current level of transparency and options?
AI robots' decision-making mechanisms	How can researchers consider AI robots' decision-making mechanisms from an extended ethical implications perspective? How do AI robots' different levels of decision-making capabilities affect the locus of ethical responsibility pertaining to critical incidents (e.g., in case of autonomous vehicles)?
AI robots' learning mechanisms	How can researchers consider AI robots' learning mechanisms from an extended ethical implications perspective? What are the actions/ways of use that can improve learning in an ethical way? Researchers could apply the presented new framework of illegal, immoral, permissible, and supererogatory use of AI robots to study AI robots' learning processes This could inform action on how relevant learning processes are regulated, for instance, by avoiding the learning of mocking others or deception
AI robots' near-humanness	How does the increasing sophistication affect humans' approaches towards AI robots? Although currently the lack of sentience and consciousness (Hildt, 2019) are generally accepted traits of AI robots as non-human beings, we expect that their humanness would further increase in the future. Thus, like service animals, AI robots may be subject to special attention and care. Future research could explore how this dynamic relationship between humans and AI robots evolves

and thus suggest further empirical investigation, the ethical nature of our study justifies using extant cases. Ethicists differ considerably in their approaches to using empirical data. Theorists argue that the empirical branch of business ethics lacks thorough theoretical grounding due to the focus on data rather than reasoning (Doorn, 2010). However, pragmatically oriented ethicists argue that empirical descriptions, in the form of considering the pragmatic conditions, are vital parts of applied ethics and that there is a need to fill the gap between ethical principles and guidance of action with empirical considerations (Birnbacher, 1999). This is to improve the argumentation and to increase the research applicability into ethics in support of everyday judgments and decisions. However, even among the pragmatically oriented ethicists the general rational is to bring in empirical insights through various ways and incorporate them in a reflective manner (Van Thiel & Van Delden, 2010) rather than necessarily having to collect primary data—for instance, through interviewing patients and medical professionals who use AI robots for increased healthcare outcomes—to engage with ethical thought. Although we center attention on AI

robots in this study, which simultaneously represents a focus and a limitation, researchers potentially can extend the presented new framework to the ethical investigation of a wider group or specific groups of AI applications, such as AI robots' use in healthcare, educational settings, or the energy sector.

Table 4 identifies several potential future research topics, and related research questions. The illegal/immoral/permissible and supererogatory normative ethical categories, for instance, are worthy of further research. This future investigation could include the study of moral responsibility, for instance, by assessing how responsibility evolves across different cases and different levels of automation. Learning and decision-making mechanisms are highly relevant and so future research could explore how organizations should manage them in an ethically responsible manner. At a more applied level, fellow researchers could further refine ethical implications specific to using AI support systems in which AI robots are embedded.

Declarations

Conflict of interest The authors whose names are listed above certify that they have no affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; grants; participation in speakers' bureaus, membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal/professional relationships, affiliations, knowledge, or beliefs) in the subject matter or materials discussed in this manuscript.

References

Admoni, H., & Scassellati, B. (2017). Social eye gaze in human-robot interaction: A review. *Journal of Human-Robot Interaction, 6*(1), 25–63.

Aleksander, I. (2017). Partners of humans: A realistic assessment of the role of robots in the foreseeable future. *Journal of Information Technology, 32*(1), 1–9.

Alexander, L., & Moore, M. (2007). *Deontological ethics*. Stanford: Stanford Encyclopedia of Philosophy.

Allen, C., & Wallach, W. (2014). Moral machines: Contradiction in terms or abdication of human responsibility? In P. Lin, K. Abney, & G. A. Bekey (Eds.), *Robot ethics: the ethical and social implications of robotics* (pp. 55–68). MIT Press.

Alles, M., & Gray, G. L. (2020). Will the medium become the message? A framework for understanding the coming automation of the audit process. *Journal of Information Systems, 34*(2), 109–130.

Arrow, K. (1973). Social responsibility and economic efficiency. *Public Policy, 21*, 303–317.

Balakrishnan, J., Malhorta, A., & Falkenberg, L. (2017). Multi-level corporate responsibility: A comparison of Gandhi's trusteeship with stakeholder and stewardship frameworks. *Journal of Business Ethics, 141*, 133–150.

Bardy, R., Drew, S., & Kennedy, T. F. (2012). Foreign investment and ethics: How to contribute to social responsibility by doing business in less-developed countries. *Journal of Business Ethics, 106*(3), 267–282.

Baskerville, R., Myers, M., & Yoo, Y. (2020). Digital first: The ontological reversal and new challenges for IS. *MIS Quarterly, 44*(2), 509–523.

Benlian, A., Klumpe, J., & Hinz, O. (2019). Mitigating the intrusive effects of smart home assistants by using anthropomorphic design features: A multimethod investigation. *Information Systems Journal*. https://doi.org/10.1111/isj.12243

Bench-Capon, T. J. (2020). Ethical approaches and autonomous systems. *Artificial Intelligence, 281*, 1–28.

Bera, P., Soffer, P., & Parsons, J. (2019). Using eye tracking to expose cognitive processes in understanding conceptual models. *MIS Quarterly, 43*(4), 1105–1126.

Beu, D., & Buckley, M. R. (2001). The hypothesized relationship between accountability and ethical behavior. *Journal of Business Ethics, 34*(1), 57–73.

Bilgeri, D., Fleisch, E., Gebauer, H., & Wortmann, F. (2019). Driving process innovation with IoT field data. *MIS Quarterly Executive, 18*(3), 191–207.

Birnbacher, D. (1999). Ethics and social science: Which kind of cooperation? *Ethical Theory and Moral Practice, 2*(4), 319–336.

Bommer, M., Gratto, C., Gravander, J., & Tuttle, M. (1987). A behavioral model of ethical and unethical decision making. *Journal of Business Ethics, 6*(4), 265–280.

Broekens, J., Heerink, M., & Rosendal, H. (2009). Assistive social robots in elderly care: A review. *Gerontechnology, 8*(2), 94–103.

Bruder, I. (2020). A social mission is not enough: Reflecting the normative foundations of social entrepreneurship. *Journal of Business Ethics*. https://doi.org/10.1111/isj.12243

Buhmann, A., Paßmann, J., & Fieseler, C. (2019). Managing algorithmic accountability: Balancing reputational concerns, engagement strategies, and the potential of rational discourse. *Journal of Business Ethics*. https://doi.org/10.1007/s10551-019-04226-4

Byrne, E. F. (2018). Making drones to kill civilians: Is it ethical? *Journal of Business Ethics, 147*(1), 81–93.

Calo, R. (2017). Artificial Intelligence policy: A primer and roadmap. *U. c. Davis Law Review, 51*(2), 399–436.

Čaić, M., Odekerken-Schröder, G., & Mahr, D. (2018). Service robots: Value co-creation and co-destruction in elderly care networks. *Journal of Service Management, 29*(2), 178–205.

Constantinescu, M., & Kaptein, M. (2015). Mutually enhancing responsibility: A theoretical exploration of the interaction mechanisms between individual and corporate moral responsibility. *Journal of Business Ethics, 129*(2), 325–339.

Choudhary, S., Arbat, H., & Patkar, U. (2016). An Innovative study on artificial intelligence and robotics. *International Journal of Innovative Research in Computer and Communication Engineering*. https://doi.org/10.15680/IJIRCCE.2016.0403062

Ciborra, C., & Willcocks, L. (2006). The mind or the heart? It depends on the (definition of) situation. *Journal of Information Technology, 21*(3), 129–139.

Cropanzano, R., Goldman, B., & Folger, R. (2013). Deontic justice: The role of moral principles in workplace fairness. *Journal of Organizational Behavior, 24*, 1019–1024.
Doorn, N. (2010). Applying Rawlsian approaches to resolve ethical issues: Inventory and setting of a research agenda. *Journal of Business Ethics, 91*(1), 127.
Dormehl, L. (2018). *Thinking machines: The inside story of Artificial Intelligence and our race*. Ebury Publishing.
Driver, J. (1992). The supererogatory. *Australasian Journal of Philosophy, 70*(3), 286–295.
Etzioni, A., & Etzioni, O. (2017). Incorporating ethics into artificial intelligence. *The Journal of Ethics, 21*(4), 403–418.
Fahlquist, J. N. (2009). Moral responsibility for environmental problems—Individual or institutional? *Journal of Agricultural and Environmental Ethics, 22*(2), 109–124.
Fleming, P. (2019). Robots and organization studies: Why robots might not want to steal your job. *Organization Studies, 40*(1), 23–38.
François, A., Bayle, E., & Gond, J. P. (2019). A multilevel analysis of implicit and explicit CSR in French and UK professional sport. *European Sport Management Quarterly, 19*(1), 15–37.
Freeman, R. E. (1994). The politics of stakeholder theory: Some future directions. *Business Ethics Quarterly, 4*(4), 409–421.
Friedman, M. (1970). The social responsibility of the business is to increase its profits. *New York Times Magazine*, 13 September 1970.
Gunkel, D. J. (2012). *The machine question: Critical perspectives on AI, robots, and ethics*. MIT Press.
Hasnas, J. (1998). The normative theories of business ethics: A guide for the perplexed. *Business Ethics Quarterly, 8*(1), 19–42.
Hassan, N. R., Mingers, J., & Stahl, B. (2018). Philosophy and information systems: Where are we and where should we go?. *European Journal of Information Systems, 27*(3), 263–277.
Heath, J. (2007). An adversarial ethic for business: Or when Sun-Tzu met the stakeholder. *Journal of Business Ethics, 72*(4), 359–374.
Heath, J. (2008). *Following the rules: Practical reasoning and deontic constraint*. OUP.
Heath, J. (2011). Three normative models of the welfare state. *Public Reason, 3*(2), 13–43.
Heath, J. (2014). *Morality, competition, and the firm: The market failures approach to business ethics*. Oxford University Press.
Hildt, E. (2019). Artificial Intelligence: Does Consciousness Matter? *Frontiers in Psychology, 10*, 1535.
Hitt, M. A., Beamish, P. W., Jackson, S. E., & Mathieu, J. E. (2007). Building theoretical and empirical bridges across levels: Multilevel research in management. *Academy of Management Journal, 50*(6), 1385–1399.
Hu, Z., Liu, B., & Zhao, Y. (2018). Agricultural robot for intelligent detection of Pyralidae insects. In agricultural robots-fundamentals and applications. *IntechOpen*. https://doi.org/10.5772/intechopen.79460
Huang, M. H., & Rust, R. T. (2011). Sustainability and consumption. *Journal of the Academy of Marketing Science, 39*(1), 40–54.
Hursthouse, R. (1999). *On virtue ethics*. Oxford University Press.
Jiang, J., & Cameron, A. F. (2020). IT-enabled self-monitoring for chronic disease self-management: An interdisciplinary review. *MIS Quarterly, 44*(1), 451–508.
Jones, T. M. (1991). Ethical decision making by individuals in organizations: An issue-contingent model. *Academy of Management Review, 16*(2), 366–395.
Johnson, M. (1994). *Moral imagination: Implications of cognitive science for ethics*. University of Chicago Press.
Johnson, D. G. (2015). Technology with no human responsibility? *Journal of Business Ethics, 127*(4), 707–715.
Kagan, S. (2018). *Normative ethics*. Routledge.
Kamishima, Y., Gremmen, B., & Akizawa, H. (2018). Can merging a capability approach with effectual processes help us define a permissible action range for AI robotics entrepreneurship? *Philosophy of Management, 17*(1), 97–113.
Kaplan, A., & Haenlein, M. (2020). Rulers of the world, unite! The challenges and opportunities of artificial intelligence. *Business Horizons, 63*(1), 37–50.
Katz, P. S. (2013). Expert robot: Using artificial intelligence to assist judges in admitting scientific expert testimony. *Albany Law Journal of Science and Technology, 24*, 1–44.
King, B. F., Jr. (2017). Guest editorial: Discovery and artificial intelligence. *American Journal of Roentgenology, 209*(6), 1189–1190.
Khalil, O. E. (1993). Artificial decision-making and artificial ethics: A management concern. *Journal of Business Ethics, 12*(4), 313–321.
Koslowski, P. (2001). Economics and ethics I. *Principles of ethical economy* (pp. 38–80). Springer.
Leventi, N., Yanakieva, A., Vodenicharova, A., & Deliverska, M. (2017). Bioethics–roboethics: Social and ethical implications of sciences development. *American Journal of Engineering Research, 6*(12), 340–343.
Lin, P., Abney, K., & Bekey, G. A. (2014). *Robot ethics: The ethical and social implications of robotics*. MIT Press.
Lobschat, L., Mueller, B., Eggers, F., Brandimarte, L., Diefenbach, S., Kroschke, M., & Wirtz, J. (2021). Corporate digital responsibility. *Journal of Business Research, 122*, 875–888.
Longoni, A., & Cagliano, R. (2018). Sustainable innovativeness and the triple bottom line: The role of organizational time perspective. *Journal of Business Ethics, 151*(4), 1097–1120.
Martin, K. (2019). Ethical implications and accountability of algorithms. *Journal of Business Ethics, 160*(4), 835–850.
Ma, T., & McGroarty, F. (2017). Social Machines: How recent technological advances have aided financialisation. *Journal of Information Technology*, *32*(3), 234–250.
Mazutis, D. (2014). Supererogation: Beyond positive deviance and corporate social responsibility. *Journal of Business Ethics, 119*(4), 517–528.
Miller, A. (2003). *An introduction to contemporary metaethics*. Blackwell.
Moor, J. H. (2006). The nature, importance, and difficulty of machine ethics. *IEEE Intelligent Systems, 21*(4), 18–21.
Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.
Norman, W. (2011). Business ethics as self-regulation: Why principles that ground regulations should be used to ground beyond-compliance norms as well. *Journal of Business Ethics, 102*(1), 43–57.
Peterson, M. (2013). *The dimensions of consequentialism: Ethics, equality and risk*. Cambridge University Press.
Reuters. (2020). *'Outrageous' that data deleted from main UK police computer database*, PM Johnson says, 20.01.2020, www.reuters.com
Russell, S., Hauert, S., Altman, R., & Veloso, M. (2015). Ethics of artificial intelligence. *Nature, 521*(7553), 415–416.
Soares, C. (2003). Corporate versus individual moral responsibility. *Journal of Business Ethics, 46*(2), 143–150.
Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review, 20*(3), 571–610.
Tahir, A. M., Naselli, G. A., & Zoppi, M. (2018). Soft robotics: A solid prospect for robotizing the natural organisms. *Advances in Robotics Research, 2*(1), 69–97.
Torresen, J. (2018). A review of future and ethical perspectives of robotics and AI. *Frontiers in Robotics and AI, 4*, 75–85.
Trevino, L. K., & Brown, M. E. (2004). Managing to be ethical: Debunking five business ethics myths. *Academy of Management Perspectives, 18*(2), 69–81.
Tsoukas, H. (2020). Leadership, the American Academy of Management, and President Trump's travel ban: A case study in moral imagination. *Journal of Business Ethics, 163*(1), 1–10.

Van Thiel, G., & Van Delden, J. (2010). Reflective equilibrium as a normative empirical model. *Ethical Perspectives, 17*(2), 183–202.
Veruggio, G., Operto, F., & Bekey, G. (2016). Roboethics: Social and ethical implications. *Springer Handbook of robotics* (pp. 2135–2160). Springer.
Vogel, D. (1992). The globalization of business ethics: Why America remains distinctive. *California Management Review, 35*(1), 30–49.
Wang, X., Tajvidi, M., Lin, X., & Hajli, N. (2020). Towards an ethical and trustworthy social commerce community for brand value co-creation: A trust-commitment perspective. *Journal of Business Ethics, 167*(1), 137–152.
Westerlund, M. (2020). An ethical framework for smart robots. *Technology Innovation Management Review, 10*(1), 35–44.
Wilson, H. R., & Series, F. L. (2002). The constantly rising ethics bar. In: *Presentation to the Canadian Centre for Ethics and Corporate Policy*, Toronto, 7 November 2002.
Wirtz, J., Patterson, P. G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S., & Martins, A. (2018). Brave new world: Service robots in the frontline. *Journal of Service Management, 29*(5), 907–931.
Wright, S. A., & Schultz, A. E. (2018). The rising tide of artificial intelligence and business automation: Developing an ethical framework. *Business Horizons, 61*(6), 823–832.
Yoon, S. N., & Lee, D. (2019). Artificial intelligence and robots in healthcare: What are the success factors for technology-based service encounters? *International Journal of Healthcare Management, 12*(3), 218–225.
Young, C. (2015). *Joseph Heath's Morality, competition, and the firm: The market failures approach to business ethics*. Oxford University Press.
Young, S., & Marais, M. (2012). A multi-level perspective of CSR reporting: The implications of national institutions and industry risk characteristics. *Corporate Governance: An International Review, 20*(5), 432–450.
Zieba, S., Polet, P., & Vanderhaegen, F. (2011). Using adjustable autonomy and human–machine cooperation to make a human–machine system resilient—Application to a ground robotic system. *Information Sciences, 181*(3), 379–397.
Zsolnai, L. (2006). *Interdisciplinary yearbook of business ethics* (pp. 53–86). GSE & Peter Lang Publishing Group.