Moral Approaches to AI: Missing power and marginalized stakeholders

Carolina Villegas-Galaviz
Kirsten Martin

The introduction of AI to augment business decisions has strained the standard ethical approaches in business ethics, where the firm is to focus on the interests of stakeholders (SHs). Unique attributes of AI and AI research – reinforcing systems of power, surreptitious yet pervasive data collection, and marginalizing vulnerable SHs – can be better addressed through specific normative approaches that raise the voice of the marginalized SHs either by focusing on the power dynamics of the larger socio-technical system or by prioritizing the relationships between actors and their unique vulnerabilities. The goal of this article is to examine the prominent moral approaches to the ethics of Artificial Intelligence (AI) in business ethics, identify the strengths and limitations of each approach to the field, and propose normative approaches focused on power and vulnerable SHs as needed within the examination of AI within business ethics.

Moral Approaches to AI: Missing power and marginalized stakeholders

Introduction

The use of algorithms and data have frequently come to public scrutiny with scandals of power abuse or violations of rights, just as privacy. For example, Pasco schools and the sheriff's office uses predictive analytics to identify future criminals from the school's rosters (Bedi and McGrory, 2020), organizations are using artificial intelligence (AI)[i] for hiring and promotion decisions with discriminatory results (Ajunwa, 2019), and Facebook employs content recommendation algorithms that promote hate groups and discriminated against Black users (Hasan et al. 2022; Dwoskin et al. 2021).

AI development is faster than the associated ethical deliberation, and the understanding of ethical issues for those who develop and deploy AI many times had come after the harm has been done (Martin and Freeman, 2004). AI can be related to a pejorative feeling within society and directly connected with risk (Araujo et al. 2020). However, stopping the damage is not always easy or quick. While challenging to anticipate and prepare for the unknown, concepts and ethical approaches can help ameliorate the harm created by AI.

While many other fields study the ethics of AI, business ethics is in a unique position to both normatively examine AI as well as the associated responsibility of the firm. And within the last few years, the use of AI for pricing (Seele et al. 2021), behavioral tracking (Steinberg, 2021), social media addiction (Bhargava and Velasquez, 2021), or gamification (Kim, 2018), has been the subject of ethical examination within business ethics thus bringing important attention to the firm decision and their moral implications.

In this paper, we analyze the prominent normative approaches to AI, identify the associated questions those approaches seek to address, and the limitations that each one may encounter. Unique attributes of AI and AI research – reinforcing systems of power, surreptitious, pervasive data collection (Shilton et al. 2021), marginalizing vulnerable stakeholders (SHs)– can be better addressed through specific normative approaches that raise the voice of the marginalized SHs either by focusing on the power dynamics of the larger socio-technical system or by prioritizing the relationships between actors and their unique vulnerabilities. Critical approaches and the ethics of care offer a unique approach to critically examining AI within business ethics. As such, this paper contributes to

business ethics scholarship by offering novel normative approaches to the study of AI and its moral implications. These approaches center marginalized SHs, discussions around vulnerabilities and relationships, and the power dynamics of the current socio-technical systems for AI.

We illustrate our study with one specific example, the case of how Amazon uses algorithms to rate its drivers, provide feedback, and, if deemed necessary, fire drivers by email (Soper, 2021). While Amazon's drivers can digitally see their rating of *Fantastic*, *Great*, *Fair*, or *At Risk*, drivers only receive automated feedback. Any termination is also only done through an email. We use this example to analyze how each normative approach addresses the ethical implications of AI.

For teaching the ethics of AI within business schools or corporations, this paper offers pragmatic questions to ask in the design, development, and use of AI. This would serve as a roadmap for people designing, developing, and using AI programs to question the moral implications in their part of the process.

Dominant, Normative Approaches to AI Ethics

Contrary to claims that AI, including machine learning, data analytics, and other types of computer programs, is objective or neutral, AI embodies the value-laden decisions of programmers and has moral implications when in use. In other words, decisions using AI models can diminish the rights of individuals, harm SHs, violate rules, norms, and laws, as well as unjustly distribute social goods (Martin, 2019a). How to assess those moral implications as being ethical or unethical has relied on four dominant approaches to AI ethics: deontology, justice/fairness, virtue ethics, and autonomy and responsibility approaches.

Deontological or Principle-based Ethics

Deontological ethics refers to normative ethical approaches based on duties. In brief, according to deontologists, the moral rightness of an action depends on its accordance with the agent's obligations. An act is good if the person fulfills his or her duty. Deontology is frequently explained as opposed to utilitarian ethics and consequentialism, where the outcome determines the rightness of actions. Within business ethics, deontology is usually related to Kantian ethics and frequently portrayed as excessively formalistic, although Kant himself talked about virtue, character, and the teleological essence of actions (Dierksmeier, 2013). However, the fulfillment of duties goes in line with following some ethical principles that one should apply independently of own preferences.

In the face of a new ethical scenario, guidelines or codes of conduct appear as a first way to secure the ground. Those principles set out the duties that each person must fulfill and establish the limits that should not be crossed. In AI ethics, the first impulse has been towards the search for principles to guide developers and users in unknown terrain. A decade ago, the field was "concerned with giving machines ethical principles, or a procedure for discovering a way to resolve the ethical dilemmas" (Anderson and Anderson, 2011).

Several moral guidelines on AI have been proposed in the search for moral principles that guide machine ethics. By 2019 there were at least 84 guidelines for ethical AI (Jobin et al. 2019). Principles have come from academia, governments, private institutions, non-

profit organizations, and professional associations. Some of these rules or guidelines refer to the traditional principles approach of bioethics: beneficence, non-maleficence, justice, and autonomy (Mittelstadt, 2019; Floridi et al. 2018; Lepri et al. 2018). Others focus on the specific nature of AI and propose principles referring to transparency, responsibility, privacy, freedom, trust, sustainability, and solidarity (Jobin et al. 2019).

Most of the guidelines allude to what they identify as universal principles for all ethical agents, which imply the respect of agents other than the self. Also, some are adaptations of preestablished ethical norms from other disciplines (Mittlestadt, 2019) or in other contexts (Vidgen et al. 2020).

Hence, following a deontological approach within AI, one should ask: What are the duties and responsibilities for this program? Am I fulfilling my duties in designing, developing, or deploying this algorithm? How can I ensure I fulfil the duty for transparency and beneficence?

While principles appear to be necessary for ethics, principles are not sufficient and "alone cannot guarantee ethical AI" (Mittelstadt, 2019). In AI ethics, although there is convergence around certain ethical principles, there are fundamental divergences about "(1) how ethical principles are interpreted; (2) why they are deemed important; (3) what issue, domain or actors they pertain to; and (4) how they should be implemented" (Jobin et al. 2019). Under all this, the problem of power balance appears in the interpretations and definitions of most ethical concepts. Authors refer to an underrepresentation of geographic areas (Jobin et al. 2019) and to how cultural differences tend to be ignored, and marginalized ethical traditions (as African Ethics) are not referenced, but there is a hegemony of western approaches (Segun, 2021). In addition, strategies to improve the effective adoption of AI principles have already been suggested, which imply components such as training, having an ethics office(r), or reporting mechanisms (Kelley, 2022).

In the example of Amazon, according to deontology, those who develop and deploy the algorithm should ask: what are the duties and responsibilities in the design, development, and use of AI in evaluating drivers? Or (for example) is the model transparent and explainable? The Amazon example demonstrates the limitations of a deontological approach. There is no stated prohibition on firing someone via email, no breach of duty. However, those who designed the algorithm and Amazon while using it, appeared to lack the virtue of empathy. Since they do not show compassion or concern

for others (Vallor, 2016). They do not have the courtesy to fire the drivers in person and do not allow them to defend their point in the face of dismissal.

Table 1 summarizes the normative approaches to AI, where we offer an outline of the questions that each theory addresses, the contribution or what each focus offers to AI ethics, and lastly, the limitations they present.

Table 1: Summary of Normative Approaches to AI

Approach	Addresses Questions	Offers	Limitations
Dominant Normative Approaches to AI			
Principle-based	What principles should I follow when I develop or deploy AI?	Ethical codes or principles to develop and deploy AI.	The interpretation, relevance, and implementation of principles vary according to actors.
Justice	Is the AI treating people unfairly or creating unfair outcomes?	The comprehension of issues of fair distribution, rights, and equity.	The approach fails to conquer the needed change for real social justice, while focusing only on particular actors and an emphasis on disadvantages.
Virtue	Which are the moral virtues needed to behave ethically in the AI era? "How can humans hope to live well in a world made increasingly more complex and unpredictable by emerging technologies?" (Vallor, 2016).	Studies of the flourishment of humans in an uncertain future, where the uncertainty comes from the changing nature of emerging technologies.	Reliance in people good will may not be sufficient to mediate conflicts, to focus on the world around, and to pay attention to biases and imbalances in power.
Responsibility	Who should be held accountable of AI outcomes and mistakes?	The study and understanding of accountability in the development and deployment of AI.	When it fails to understand technology as value-laden and to make propositions of how to fulfill responsibilities.
Critical Normative Approaches to AI			
Critical Theories	Who is marginalized by the design of this AI program? Whose power is reinforced by the introduction of a given AI program?	To identify and critique systemic power relations with an intention to contribute to structural change and even emancipation	Over focus on power and the marginalize, when not all ethical issues are about that.
Ethics of Care	Whose voices are being silenced? Which vulnerabilities are being exploited? Is the algorithm considering context and circumstances? Are interdependent relationships considered or misused?	The understanding of AI ethics within a web of interdependent relationships, where vulnerabilities, what the other has to say, and context and circumstances play an important role for AI development and deployment.	Possible misunderstandings of the theory as altruism, feelings of pity, or something limited to feminism.

Ethics of Justice and Fairness

A common theme across justice scholarship is that an ethics of justice "places a premium on individual autonomous choice and equality" and "encompass notions of balancing rights and responsibility" (French and Weis, 2000). Within AI ethics, fairness and justice approaches deal with egalitarianism and discrimination.

The initial claim was that AI decision-making, based on quantifiable terms, could lead to more objective and, thus, more just processes. Algorithmic decision-making appeared fairer than "those made by humans who may be influenced by greed, prejudice, fatigue, or hunger" or any other feeling (Lepri et al. 2018). However, progress has gone towards determining technology as value-laden, and there is the understanding that algorithms are not neutral: data is biased, assumptions are value-laden decisions made by humans, and mistakes are unfairly distributes (Martin, 2019a). This technology can exacerbate issues regarding fair distribution, rights, and equity with the automation and acceleration of processes.

Interrelated with this appears the problem of discrimination. According to Barocas and Selbst (2016), "by definition, data mining is always a form of statistical (and therefore seemingly rational) discrimination." An algorithm learns with a set of data (input) in which it finds patterns that will later apply in new decision-making scenarios (output). The point of data mining is to provide a bolster to statistically frame individuals. When there is a new individual, algorithms confer the frame of those statistically related (which could lead to *bias*). Hence, everyone judged or determined by algorithms will always be affected by information that is not their own. This process has the potential to improperly disregard legally protected classes and lead to a *disparate impact* of big data's processes (Barocas and Selbst, 2016). Where disparate impact "refers to policies or practices that are facially neutral but have a disproportionately adverse impact on protected classes" (p. 694).

There appears a problem of data representativeness. Not only because of the possibility of overfitting and other possible manipulations of developers, but because data are reductive representations of a phenomenon with multiple possibilities and characteristics (Barocas and Selbst, 2016; Carusi, 2008; Lum, 2017). In this line appears the problem of bias, which can be *preexisting* (in society), *technical*, and *emergent* (from the context of use) (Friedman and Nissenbaum, 2016). We talk about bias only when

unfair discrimination is systematically, and it is combined with an unfair outcome (Friedman and Nissenbaum, 1996).

Many scholars have attempted to mitigate algorithmic biases and the associated injustices generally (Baer et al. 2020; Grgic-Hlaca et al. 2016; Lepri et al. 2018; Lin et al. 2020; Rahwan, 2020), and in specialized disciplines like law (Hacker, 2017) and the financial industry (Zhang and Zhou, 2019). Above all, the justice approach identifies and analyzes how algorithms serve as tools that prejudice egalitarianism and reinforce racism and discrimination while limiting possibilities for some groups of people (O'Neil, 2016). Here, Mimi Onuoha (2018) talks about *algorithm violence*, which she defines as "the violence that an algorithm or automated decision-making system inflicts by preventing people from meeting their basic needs."

Following a justice and fairness approach, those who develop and deploy AI would ask: does this outcome create disparate impact on protected classes of individuals? Is it fair to use these variables, or could these variables impact any issue of equality? Does the selected data contain any discriminatory bias? Are issues concerning egalitarianism an essential factor in the process of development and deployment of the model? Is diversity a concern within the team that develops or deploys this AI model? Are the last fortunate in society further harmed by the use of this AI program?

Fairness and justice approaches have shed much light on AI ethical issues while also having limitations. One of the main problems in finding a solution in line with fair algorithms is to conquer consensus on what it means for AI to be fair (Binns, 2018). Furthermore, since "fairness metrics which are appropriate in one context will be inappropriate in another" (Binns, 2018:), and "what constitutes fairness changes according to different worldviews" (Lepri et al. 2018): some scholars proposed that the answer would come from interdisciplinary teams working to develop fair AI (Lepri et al. 2018).

Moral and political philosophers have long been debating similar issues and concepts, but with AI, the definitions of concepts as *fairness*, *discrimination*, and *egalitarianism* take a significant new perspective. Then, AI "faces an upfront set of conceptual ethical challenges" (Binns, 2018), and some of the answers to conquer fairness in these systems will require a reexamination of the meaning of *discrimination* and *fairness*, a call for caution, and a careful application of data mining processes (Barocas and Selbst, 2016).

Another limitation of justice approaches to AI ethics is that the discourse of discrimination, rights, and fair processes fails to conquer the needed change for real social justice. The causes of this problem, among others, are the continuous emphasis on the wrong behavior of particular actors, which ignores the fact that discrimination is a social phenomenon. The development of this discourse has an exclusive focus on disadvantages, avoids propositions, and limits itself to criticism (Hoffman, 2019). Lastly, “the outsize focus on a limited set of goods downplays the role of social attitudes and background norm-setting in shaping not only people’s well-being, but our very ability to conceive and pursue particular visions of justice” (Hoffman, 2019).

In the example of Amazon, applying this approach, one should ask: does the selected data contains unfair variables that frame the drivers? Is the algorithm terminating people with discriminatory implications? Is the termination of drivers impacting specific groups of people constantly? Is the termination leading to a disparate impact? However, even though much needed, the approach does not allocate responsibilities or give concrete proposals to resolve the case and its possible ethical issues

Virtue Ethics

Virtue ethics is one of the main approaches in business ethics (Solomon, 1992; Koehn, 1995; Sison, 2015). Primarily based on the Aristotelian propositions in the *Nicomachean Ethics*, this theory is usually related to the agent's character traits, "while utilitarianism concentrates on outcomes and deontological ethics on the act itself" (Koehn, 1995). However, this should not lead to the understanding that the outcome is not essential to this theory (Koehn, 1995).

In the AI ethics field, virtue ethics appears as an approach focused on the individual rather than deontological AI ethics based on strict rules, duties, and imperatives (Hagendorff, 2020; Ananny, 2016). Hence, this research stream defends that within AI, if the predominant deontological approach leans within virtue ethics, then the AI ethics will “no longer understood as a deontologically inspired tick-box exercise, but as a project of advancing personalities, changing attitudes, strengthen responsibilities and gaining courage to refrain from certain actions, which are deemed unethical” (Hagendorff, 2020).

Within these propositions, Shannon Vallor's proposals lead the way to bring virtue ethics to answer the critical ethical questions of the current era (Vallor, 2010, 2012, 2015, 2016, 2017). The author proposes a virtue-driven approach to the ethics of emerging

technologies, such as AI, and an ethical strategy for promoting the moral character needed for the challenges of recent times. In *Technology and the virtues*, she adapted Aristotelian, Confucian, and Buddhist ethical reflections to create what she calls the *technomoral virtues* needed for the 21st century (Vallor, 2016).

The virtue approach tries to answer the question about "how can humans hope to live well in a world made increasingly more complex and unpredictable by emerging technologies?" (Vallor, 2016). The answer goes in line with how humans need to cultivate a type of moral character immersed in how technologies shape the world. This framework based on virtues and technologies is proposed to specify how humans should act to flourish in an uncertain future, where the uncertainty comes from the changing nature of emerging technologies.

Scholars have been trying to respond to fundamental questions of virtue ethics in the field of AI and emerging technologies, such as those about how humans can flourish and live a life worth well-living in the context of emerging technologies (Stahl et al. 2021; Kim and Mejia, 2019; Clark and Gevorkyan, 2020; Stahl, 2021). Also, there are some proposals of the inclusion of virtues in the design of AI models (Neubert, 2020; Gamez et al. 2020), of the virtue ethics approach as a framework for Artificial Moral Agents (AMAs) (Sison and Redín, 2021), and as a critical factor to humans as masters of AI to avoid unwitting slavish adherence to AI (Kim et al. 2021). Most of these applications refer to a *neo-Aristotelian* approach, where *neo* indicates the resolved variety of virtue ethics that rejects Aristoteles's views on women and slavery, as well as children, vulnerabilities, and dependence (Sison and Redín, 2021).

According to the virtue ethics scholarship, people within AI should question: how does this model impact the live-well of society? Does this model help to the flourishment of those who will be affected by it? Or does who will deploy them? Also, those who develop and deploy AI should ask, does this model represent my virtues and the character of a virtuous person?

Nevertheless, the approach applied in isolation may encounter some limitations. First, "conceptions of virtue and human flourishing are never universal. There have always been, and will always be, coherent accounts of the good life that cannot be reduced to or fully reconciled with other" (Vallor, 2017). Hence, this approach may sometimes need a referral to principles that delimit the prohibitions and duties of the person. This may help when the overreliance on people's goodwill (and own judgment) is not sufficient to mediate conflicts (Clifford, 2013). Also, "as moral agents, we should focus not on our

own struggles to be virtuous, but on the world around us" (Reader, 2007). This means that the approach "fails to pay sufficient attention to systemic biases and to imbalances in power" (Koehn, 1998). Therefore, claims with an axis on the marginalized and all those in need can complete and enrich this approach.

Referring to the example of Amazon, from this approach, one should ask, how does the process of algorithmically rating and terminating drivers by email impact the live-well of society? Does this process help or damage the flourishment of the drivers? Or of Amazon as a firm? Does the termination represent virtues like empathy, civility, flexibility, or magnanimity? Nonetheless, there is still a need to ask about responsibilities, duties, and biases within this process.

Responsibility

Normative approaches focused on responsibility issues examine the accountability of AI models and their impact. Within this approach, scholars research who is responsible for AI outcomes from a technical perspective and intentionality.

From a technical perspective, there is a need to identify who is responsible for mistakes and harms and to avoid the easy solution of blaming the *machine* when something goes wrong. One of the main problems in using AI systems is the so-called problem of many-hands, where many people participate in elaborating a final product or service (Villegas-Galaviz and Martin, 2022). The issue refers to the difficulty of identifying who is responsible for the outcome. Hence, "loosely, this problem may be described as the problem of attributing or allocating individual responsibility in collective settings" (van de Poel et al. 2015).

For AI, this problem entails the difficulty to identify who is responsible for the outcomes of a model, those who design, those who develop, or those who deploy the AI system? Here the progress has gone towards explaining that if experts design black-box algorithms and preclude individuals from taking responsibility in decision-making, they are accountable for the algorithm's implications in use (Martin, 2019b) and responsible for managing mistakes (Martin, 2019a). In these scenarios, social embeddedness and reflection are two tools for designing ethical algorithms and managing the inevitable mistakes of algorithms (Martin, 2019b).

There is a long discussion about the need for AMAs and to differentiate voluntary actions from machine operations (Sison and Redín, 2021). AMAs posit that there exist moral agents that are 'moral' and those that are amoral. The focus, therefore, of AMAs scholarship is to identify the attributes or possibilities of AMAs, with those AI programs not meeting that threshold being labeled amoral. The issue, of course, is that all AI is value-laden, moral, with ethical implications (Johnson, 2022; Martin, 2019a; Rudner, 1953). As Rudner rightly noted many decades ago, pretending that the design and development of technology and science is amoral just means one designs and develops technology leaving the moral decisions encoded in design unmanaged and made thoughtlessly (Rudner, 1953).

Here to shed light on this complexity of attributing responsibility, appeared the notion of *Technological Moral Action* (TMA), which combines the participation of computer system users, system designers (developers, programmers, and testers), and computer systems (hardware and software) (Johnson and Powers, 2005). The notion of TMA adds the idea that to ascribe responsibility, the part played by technology should be considered. This means that looking only at humans' free and intended actions is not enough. The notion is a try to introduce artifacts into the sights of moral responsibility and avoid the understanding of technology and its outcomes as natural phenomena. "Moral responsibility is focused on behavior that is freely chosen, and in TMA the user and the artifact-maker have acted freely and could have done otherwise. Because the artifact is freely made, it could be otherwise" (Johnson and Powers, 2005).

The quid is that even though computer systems are moral entities, they are not moral agents since they are components in moral actions (Johnson, 2015). AI systems could not be considered moral agents because of their lack of mental states and intending's to act, which are particular of agent's freedom (Johnson, 2006). However, AI systems are not neutral because they are "intentionally created and used forms of intentionality and efficacy" (Johnson, 2006): then, they should be taken as part of the moral world because of their effects and what they are and do.

The idea is that technological development sometimes is seen as logically composed with an inevitable conclusion, while it is multidirectional and contingent. Hence, the appearance of a responsibility gap, or the supposition that in certain scenarios no one is really responsible for technology impacts, depends on human choices and not on the complexity of artificial agents. Humans can decide to create technologies with no human responsibility, but that could be a choice, not a result of technology's nature (Johnson,

2015; see also Sison and Redín, 2021). "Speculations about a responsibility gap misrepresent the situation and are based on false assumptions about technological development and about responsibility" (Johnson, 2015). Any claims of the responsibility gap, in other words, is constructed by ignoring the humans responsible for the value-laden design decisions and moral implications of use.

From a responsibility approach, those who design, develop, and deploy AI should ask who should be held accountable for AI outcomes? Also, they should critically examine their part in the process and the implications of each of their actions.

Still, the responsibility approach to AI ethics may encounter some limitations. Some struggle acknowledging the value-laden biases of technology – including algorithms – while preserving the ability of humans to control the design, development, and deployment of technology.[ii] Only by acknowledging the value-laden biases of algorithms can we begin to ask how companies inscribed those biases during design and development (Martin, 2022b). Unfortunately, for some claiming that technology or AI has moral agency necessitates making technological imperative arguments – framing algorithms as evolving under their own inertia, providing more efficient, accurate decisions, and outside the realm of interrogation. In their search for responsibility, *technological determinists* see technology as to 'blame' for the outcome. Johnson (2006) provides an excellent example of how to acknowledge the moral implications of AI as an actor without attributing moral agency to an artifact, yet many fall victim to this mistake in their effort to identify AI as *doing* immoral things.

In AI ethics, is not enough to allocate responsibilities, there is a need of another approach. The focus on responsibility alone finds its limit in the identification of how to do things right or how to fulfill responsibilities.

In the example of Amazon, within this approach, one should ask: who is responsible for the harm in the termination of the drivers? Also, who is responsible for managing mistakes in wrong dismissals? Here, other needed questions as how to develop a model which helps to the flourishment of those impacted by it, are the focus of other approaches, as virtue ethics.

Normative Theories about Power and the Vulnerable

AI is increasingly implemented within systems of control and power, where users are rendered more vulnerable through the implementation of AI programs. As Ari Waldman correctly states, "using algorithms to make commercial and social decisions is really a story about power, the people who have it, and how it affects the rest of us" (Waldman, 2019, p. 615). While all "data are a form of power" (Iliadis and Russo, 2016), predictive analytics are used to "impose order, equilibrium, and stability to the active, fluid, messy, and unpredictable nature of human behaviour and the social world at large." (Birhane, 2021). And the marketplace is "inherently political with social and structural relations that connect to inequalities," which include ethnicity, race, gender, sexual orientation, religion, and physical disability (Henderson and Williams, 2013; Poole et al. 2021). Within the critical examination of the Big Tech, previous research has focused on the damaging influence of corporations on the direction of AI ethics research (Abdalla and Abdalla, 2021), the power of the corporation over data and privacy (Waldman, 2021), and powerful corporations prioritizing efficiency and freedom for some over other values (Cohen, 2019; Waldman, 2019).

In addition, while defining big data and big data ethics around the 4 Vs is popular to emphasize the bigness of the new data sets, big data sets have been in use for decades. As noted by Shilton et al, "the notable change is not the 'bigness' of digital datasets, but the ubiquitous nature of the data sources and collection methods" that allow firms develop AI programs to categorize and predict individuals using these "multiple, partial, and disconnected datasets" (Shilton et al. 2021). This introduces distance between the firms developing the AI program and users who are unaware of the value-laden decisions being made with their data and about them.

Finally, the subjects of an AI program – used in the training data and subject to the decisions of the AI program – do not have a voluntary, mutually beneficial relationship with the firm as is normally assumed (Freeman et al. 2020). Instead, subjects of the AI program are legitimate but marginalized SHs by being not only the most impacted, but also the SHs without voice or power in the design and implementation of AI models.

These unique attributes of AI and AI research – reinforcing systems of power, surreptitious, pervasive data collection, and marginalizing vulnerable SHs – can be better addressed through specific normative approaches that raise the voice of the marginalized

SHs either by focusing on the power dynamics of the larger socio-technical system or by prioritizing the relationships between actors and their unique vulnerabilities.

Critical Approaches

Critical theoretical approaches maintain a healthy skepticism towards any assumptions of neutrality or objectivity and contextualize situations in a way that accounts for the influence of different actors – currently and historically. Importantly, critical theoretical approaches seek to identify and critique systemic power relations with an intention to contribute to structural change and even emancipation (Poole et al. 2021; Stahl, 2021).

Taking a critical approach has been used throughout the examination of AI. For example, scholars examine whether technology is helping only those with power and advantage (Mohammad, 2021) or who benefits from making predictions with AI (Kerr and Earle, 2013; Martin, 2022b). In addition, AI can be used to further disenfranchise people in poverty (Eubanks, 2018) or reinforce systemic racism (Benjamin, 2019) and misogyny (D'Ignazio and Klein, 2020) and disproportionately impact LGBTQ+ (Waldman, 2019). Even more generally, we see this critical lens being used to highlight when privacy violations harm those who are marginalized (Skinner-Thompson, 2020) or are victims of nonconsensual pornography (Citron and Franks, 2014; Keats Citron, 2018), and even the use of algorithms to undermine due process rights of individuals (Citron, 2007).

The "emancipatory intention of critical research" (Stahl, 2021) works to "demystify power struggles and support efforts to dismantle entrenched hierarchical marketplace dynamics" (Poole et al. 2021). A critical examination will question the power dynamics behind the decision to choose one alternative over other options. This explicit lens of power – who has it, who benefits from the decisions made, who is harmed by the decisions made, and how the decision to benefit certain actors and punish others fit within the existing power structure – would be turned to the design decision of AI.

Critical approaches have limitations. For example, not all ethical issues of AI center power and the marginalized. One can have an AI program that breaks rules or is unfair without the impact falling disproportionately on the less powerful.

Discussing Amazon's algorithm, from a critical approach would focus more on the existing power structures and how the program reinforces the powerful and marginalizes

the vulnerable. Which are the existing power dynamics between Amazon and the drivers? How does the design and implementation further exacerbate the power of Amazon and further marginalize the drivers? The approach would focus more on the due process rights afforded to the subjects, in addition, the goal would not only be to not exacerbate existing power imbalances but to provide an emancipatory lift to those currently being marginalized (the drivers). The framing of the problem and the possible solutions differ under the critical approach.

Ethics of Care

The ethics of care appeared as a moral framework in the XX Century. Carol Gilligan first mentioned the notion in her book *In a different voice* (1982). This approach emerged as a response to the reduction of morality to formal rationality and to a dialogue between principles and rights.

> This conception of morality as concerned with the activity of care center moral development around the understanding of responsibility and relationships, just as the conception of morality as fairness ties moral development to the understanding of rights and rules. (Gilligan, 1982)

Carol Gilligan noticed that in the studies of his advisor, Lawrence Kohlberg, about six stages of moral development, women were not considered. Kohlberg (1981), a proponent of justice approaches, based his theory on a study of eighty-four boys. Hence, when his theory was applied to the groups excluded from his original sample, these groups hardly reached the higher states of moral maturity (Gilligan, 1982). Gilligan noted that girls and women seem to stick to the third stage, when morality is conceived in interpersonal terms. Here, from the focus of an ethics of care, the problem is to not listen to the different voices, basing morality on the judgment of a few, or ignoring and rejecting opinions that are less valid because they are minority (or vulnerable). This approach originated to give women a voice, which is why the ethics of care is usually related to feminism. However, the ethics of care is not only about feminism. It has its roots in a feminist dialogue but extends to a broader spectrum of social, political, and economic applications (Villegas-Galaviz, 2022a; see also French and Weis, 2000).

Almost four decades have passed since the coined of the term. There is a broader understanding of the designations and implications of *care* within ethics (Held, 2006).

The approach has developed to a more rigorous definition based on the study of different disciplines such as moral philosophy (Held, 2006; Baier, 1985), bioethics (Harbinson, 1992; Gillon, 1992), psychology (Gilligan, 1982), political theory (Tronto, 1993; Engster, 2007), education (Noddings, 1984; 2013), and business (Hamington and Sander-Staudt, 2011).

Although there is debate regarding presenting a concrete definition (Held, 2006). Scholars in the ethics of care coincide in addressing the same concepts, questioning the same things, and approaching dilemmas from the same perspective. The literature presents the ethics of care as a relational approach, where interdependent relationships play a crucial role in ethical decision-making, in contrast to the individual approach addressed by Western propositions (Segun, 2021). Also, the ethics of care appears as a contextualized moral theory, with a specific concern to protect the marginalized, avoid harm, and advocate for the non-exploitation of people's vulnerabilities. The focus of this moral approach is to hear everyone's voice and to defend those whose voices are being silenced.

In line with delineating the scope of the ethics of care, Daniel Engster developed a definition of the notion of *care* within the ethics of care:

> Everything we do directly to help individuals to meet their vital biological need, develop or maintain their basic capabilities, and avoid or alleviate unnecessary or unwanted pain and suffering, so that they can survive, develop, and function in society. [And something that should be done] in an attentive, responsive, and respectful manner. (Engster, 2007)

Based on his delineation of what care is and in dialogue with SHs theory, he proposed a definition of the ethics of care as:

> A theory that associates moral action with meeting the needs, fostering the capabilities, and alleviating the pain and suffering of individuals in attentive, responsive, and respectful ways. (Engster, 2011)

The ethics of care has been applied to different fields of technology. Most of these works refer to care-robots (Santoni de Sio and van Wynsbergue, 2016; van Wynsbergue, 2016). Also, to engage "with discussions in science and technology studies (STS) that address the 'more than human worlds' of sociotechnical assemblages and objects as lively politically charged 'things'" and posthumanism (de la Bellacasa, 2017; see also 2011).

Moreover, the theory has also been proposed for engineering to create awareness of ethical decision making and the understanding of the *other*, and to include "the need to design technologies, goods, and services for people who are not engineers and who are also different from them on other characteristics such as gender, race, and disability" (Hersh, 2016).

The ethics of care can help bring out neglected things in the study of science and technology (de la Bellacasa, 2011). Within technoscience, the ethics of care serve as a critical approach to emphasize responsiveness and to add the intention of respect and engagement with those affected by technology. There, the theory "connotes attention and worry for those who can be harmed by an assemblage but whose voices are less valued, as are their concerns and need for care" (de la Bellacasa, 2011).

This approach has also been applied to business since the 1990s (Melé, 2014; see Hamington and Sander-Staudt, 2011). Scholars addressed the ethics of care to shed light on topics such as crisis management (Simola, 2003; Sandin, 2009), leadership (Ciulla, 2009), or consumption (Shaw et al. 2016). Also, this approach has been proposed as a moral framework for SHs theory (Wicks et al. 1994; Burton and Dunn, 1996; Engster, 2011). Here the ethics of care appear as an adequate proposal where the interests and needs of the marginalized SHs are not considered. Also, its critical approach as a contextualized moral theory offers a unique point of view for unforeseen or unintended consequences (Koehn, 2011).

Bringing together the propositions of the ethics of care in business and technology in general, we propose to address the ethics of care as moral grounding for AI ethics within business (Villegas-Galaviz, 2022b). Some authors have referred to the relevance of the ethics of care within AI ethics, making first approximations (secondary) to our objective (Rodgers and Nguyen, 2022; Telkamp and Anderson, 2022). Our proposal entails bringing the categories of ethics of care to the field of AI ethics in its applications in business (Villegas-Galaviz, 2022b; Villegas-Galaviz and Martin, 2022). Four categories of the ethics of care can help to develop and deploy an ethical AI (Villegas-Galaviz, 2022b; Villegas-Galaviz and Martin, 2022).

- The first one is *interdependent relationships*. The key here is to understand morality in a network of relationships, in interpersonal terms. From this approach, people within AI should ask, does this algorithm silence relevant interdependent relationships? Also, are interdependent relationships considered or misused? There would be essential to not take

individuals as opponents "in a contest of rights but as members of a network of relationships on whose continuation they all depend" (Gilligan, 1982).

- The second category is *context and circumstances* and refers to how the ethics of care "is a relational approach to morality that entails contextualized responsiveness to particular others" (Hamington, 2019). What the ethics of care "can mean in each situation cannot be resolved by ready-made explanations" (de la Bellacasa, 2011). Here the question appears as: is the algorithm considering context and circumstances? Also, does this algorithm eliminate context and circumstances when they can be a crucial part of a decision? Moreover, does AI open the possibility to social embeddedness?
- The third category refers to *vulnerability* and the relevance of understanding people's needs and suffering. For AI ethics, this brings out that algorithms should not prevent individuals from meeting their needs while exploiting their vulnerabilities. Here those who develop and deploy AI should ask, which vulnerabilities are being exploited? Also, does this algorithm stops the possibility of fostering the needs of protected classes or marginalized SHs?
- Lastly, the fourth category refers to *voice* or the relevance of identifying and hearing the range of voices impacted by the decision. More than a factual hearing of their voices, this refers to considering the needs of all those who are impacted by an action. From this category, people in AI should ask, whose voices are being silenced in the development and deployment of AI? Also, does this algorithm considers the needs of all the people impacted by it?

Still, like the other approaches, the ethics of care presents some limitations. As in the justice approach, the ethics of care needs to continually change its focus to offer solutions and avoid an over-emphasis on AI's disadvantages or issues. Also, as in the case of critical approaches, not all ethical issues refer to vulnerabilities or harm. Hence, there is a need for other approaches to compliment this view. Lastly, common misunderstandings of the ethics of care appear as limitations, such as thoughts of this theory as something about altruism (even it also asks for the care of oneself), feelings of pity, or something limited to feminism (Villegas-Galaviz, 2022a).

In the example of Amazon, following the four categories of the ethics of care presented above, one should ask:

- In the understanding of morality in a network of *interdependent relationships*, how can the terminations impact other members or groups of society? How are interdependent relationships being considered in this model? Does the model understand drivers as members of a community?
- Also, due to the relevance of *context and circumstances* in ethical decision-making those who develop and deploy AI should ask: are the context and circumstances of drivers considered when rating and firing? Circumstances such as weather, the state of the roads when they deliver their packages, or the holidays and their complications in finding people at home when necessary.
- Moreover, it should be asked, does the data used imply the exploitation of the drivers' *vulnerabilities*? Is certain data (personal o public) being misused? And what are the needs, issues, and interests of drivers?
- Lastly, are the *voices* of drivers and the local community being silenced? Does the algorithm consider the needs of drivers, human resources employees, and customers of Amazon? Does it consider the voices of all the people impacted by it?

Discussion and Conclusion

The purpose of this paper was to analyze the prominent normative approaches to AI, to identify the questions those formulate to AI, and the limitations that each one encounters. Our objective was to offer a roadmap for people designing, developing, and using AI, one based on questions to examine their part of the process critically.

Unique attributes of AI and AI research – reinforcing systems of power, surreptitious, pervasive data collection, marginalizing vulnerable SHs – can be better addressed through specific normative approaches that raise the voice of the marginalized SHs either by focusing on the power dynamics of the larger socio-technical system or by prioritizing the relationships between actors and their unique vulnerabilities.

Critical approaches to AI and the ethics of care are proposed as an additional approach to address whose voices are being silenced, and which vulnerabilities are being exploited?

Implications for Theory

A renewed focus on critical theories and the ethics of care in particular within the study of AI has implications not only how the field assesses the moral implications of AI, but also how the field conceptualizes corporate responsibility. First, this paper contributes to the growing field within business ethics focused on the moral examination of technology and AI in particular. While much work has been done around principles and technical definitions of fairness, the argument here is to widened the moral lenses used to examine AI to better foreground the marginalized and vulnerable SHs of the technology who are ignored in alternative approaches.

In addition, defining the moral implications for firm decisions – including design, development, and use decisions around AI – directly implicates the firm as responsible for those moral implications and broadens the field of corporate responsibility and governance. For example, when management began identifying the environmental damage of firm decisions, corporate responsibility scholarship expanded to then question what the responsibility of firms is around the environment (Driscoll and Starik, 2004; Phillips and Reichart, 2000) and critically examine who benefits from environmental initiatives (Steelman and Rivera, 2006). In a parallel manner, identifying the larger moral implications of AI design, development, and use decisions, broadens the scope of corporate responsibility scholarship. Future work could leverage corporate responsibility and governance theory to questions around AI, algorithms, and other digital technologies.

Finally, critical approaches and the ethics of care in particular bring a greater focus on the concerns and consequences of those marginalized SHs of the AI technology. For SHs theory, greater attention should be spent on those legitimate, urgent SHs with little power seen as discretionary or merely dependent SHs by Mitchell, Agle, Wood (Mitchell, Agle, Wood, 1997). As rightly noted, firms will too frequently ignore such SHs while these are legitimate SHs with real concerns and interests. In the area of AI, these are also the SHs most impacted by the design and implementation of AI. Better named marginalized SHs, these individuals and groups are both the most impacted but with the weakest voice currently in the development of AI and in our current approaches to the

normative examination of AI. Future work should better conceptualize these SHs and how to bring their concerns into the design and implementation of AI by firms.

Implications for Practice

With the introduction of AI to business and substantial investments in AI research and development, firms have focused on the search for AI ethical principles (Jobin et al. 2019). However, the effectiveness of adopting those principles has become an issue (Kelley, 2022). Also, companies address AI ethics issues according to dominant normative approaches, which present limitations when addressing the unique attributes of AI and its harms.

Our focus on the questions that each approach addresses impacts how firms comprehend matters of AI ethics, not as pre-structured guidelines but as a work in progress that needs to be questioned in every step of the design, development, and deployment of AI. Hence, it is essential to give ethical training to each individual who is part of these processes. Future work should delve into better practices to avoid the problems of the unique attributes of AI and AI research – reinforcing systems of power, surreptitious, pervasive data collection, marginalizing vulnerable SHs –. An example of best practices could be the proposition of ways to integrate empathy in the research and teaching of the design, development, and use of AI to understand the other in its circumstances and vulnerabilities.

Conclusion

We propose a broader understanding in the comprehension of how each approach presents a different and needed perspective with its own concepts. Each opens a new conversation, addresses specific problems, and asks essential questions. Here we illustrate what each theory contributes to AI ethics as a discipline. We propose the critical approaches and the ethics of care as additional approaches to the ethical examination of AI.

ENDNOTES

[i] We use the term AI to mean how algorithms "sift through data sets to identify trends and make predictions" (Martin 2019b, p 836).

[ii] Biases are value-laden design features with moral implications in use.

DECLARATIONS

Ethics approval

No ethical approval is required for this type of research.

Informed consent or Consent to publish

The authors confirm that according to the nature of this research there is no need of informed consent or consent for publications from any individual.

REFERENCES

Abdalla, M. and Abdalla, M. (2021). The grey hoodie project: Big tobacco, big tech, and the threat on academic integrity. AIES ’21: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, (pp. 287-297).

Ajunwa, I. (2019). The paradox of automation as anti-bias intervention. Cardozo L. Rev 41, 1671.

Ananny, M. (2016). Toward an ethics of algorithms: Convening, observation, probability, and timeliness. Science, Technology, & Human Values, 41(1), 93–117.

Anderson, M. and Anderson, S. L. (2011). Machine ethics. Cambridge University Press.

Araujo, T., Helberger, N., Kruikemeier, S., De Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. AI & Society 35(3), 611–623.

Baer, B. R., Gilbert, D. E., Wells, M. T. (2020). Fairness criteria through the lens of directed acyclic graphs. In M. D. Dubber, F. Pasquale & S. Das (eds.), The Oxford Handbook of Ethics of AI, (pp. 493-587). Oxford University Press, NY.

Bhargava, V. R. and Velasquez, M. (2021). Ethics of the attention economy: The problem of social media addiction. Business Ethics Quarterly 31(3), 321-359.

Baier, A. C. (1985). What do women want in a moral theory? Noûs 19(1), 53–63.

Barocas, S. and Selbst, A. D. (2016). Big data's disparate impact. California Law Review 104, 671.

Bauer, W. A. (2020). Virtuous vs. utilitarian artificial moral agents. AI & Society 35(1), 263–271.

Bedi, N. and McGrory K. (2020). Pasco's sheriff uses grades and abuse histories to label schoolchildren potential criminals. Tampa Bay Times. https://projects.tampabay.com/projects/2020/investigations/police-pasco-sheriff-targeted/school-data/ Accessed February 13 2022.

Benjamin, R. (2019). Race After Technology: Abolitionist Tools for the New Jim Code. John Wiley & Sons.

Birhane, A. (2021). The impossibility of automating ambiguity. Artificial Life 27(1), 44-61.

Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. Proceedings of the 1st Conference on Fairness, Accountability and Transparency (pp. 149–159). PMLR 81. New York University, NYC.

Burton, B. K. and Dunn, C. P. (1996). Feminist ethics as moral grounding for stakeholder theory. Business Ethics Quarterly 6(2), 133-147.

Carusi, A. (2008). Data as representation: Beyond anonymity in e-research ethics. International Journal of Internet Research 1(1), 37–65.

Ciulla, J. B. (2009). Leadership and the ethics of care. Journal of Business Ethics 88(1), 3–4.

Citron, D.K. (2007). Technological due process. Washington University Law Review, 85. 1249.

Clark, C. M., and Gevorkyan, A. V. (2020). Artificial intelligence and human flourishing. The American Journal of Economics and Sociology 79(4), 1307–1344.

Clifford, D. (2014). Limitations of virtue ethics in the social professions. Ethics and Social Welfare 8(1), 2-19.

Cohen, J.E. (2019). Between truth and power. Oxford University Press.

de La Bellacasa, M. P. (2011). Matters of care in technoscience: Assembling neglected things. Social Studies of Science 41(1), 85–106.

de La Bellacasa, M. P. (2017). Matters of care: Speculative ethics in more than human worlds (Vol. 41). U of Minnesota Press.

Dierksmeier, C. (2013). Kant on virtue. Journal of Business Ethics 113(4), 597–609.

D'Ignazio, C. and Klein, L.F. (2020). Data Feminism. MIT Press.

Driscoll, C. and Starik, M. (2004). The primordial stakeholder: Advancing the conceptual consideration of stakeholder status for the natural environment. Journal of business ethics 49(1), 55-73.

Dwoskin, E. Tiku, N., Timber, C. (2021). Facebook's race-blind practices around hate speech came at the expense of Black users, new documents show. The Washington Post: https://www.washingtonpost.com/technology/2021/11/21/facebook-algorithm-biased-race/ Accessed February 13, 2022.

Engster, D. (2007). The heart of justice. Care ethics and political theory. Oxford University Press, New York.

Engster, D. (2011). Care ethics stakeholder theory. In M. Hamington and M. Sander-Staudt (eds). Applying care ethics to business. (pp. 93-110). Springer, Oxford.

Eubanks, V. (2018). Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., Vayena, E. (2018). AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds and Machines 28(4), 689–707.

Freeman, R.E., Martin, K., Parmar, B.L. (2020). The power of AND: Responsible business without trade-offs. Columbia Business School Publishing, New York, NY.

French, W. and Weis, A. (2000). An ethics of care or an ethics of justice. Journal of Business ethics, 27, 125–136.

Friedman, B. and Nissenbaum, H. (1996). Bias in computer systems. ACM Transactions on Information Systems 14(3), 330–347.

Gamez, P., Shank, D. B., Arnold, C., North, M. (2020). Artificial virtue: The machine question and perceptions of moral character in artificial moral agents. AI & Society 35(4), 795–809.

Gilligan, C. (1982). In a different voice. Harvard University Press.

Gillon, R. (1992). Caring, men and women, nurses and doctors, and health care ethics. Journal of Medical Ethics 18(4), 171.

Grgic-Hlaca, N., Zafar, M. B., Gummadi, K. P., Weller, A. (2016). The case for process fairness in learning: Feature selection for fair decision making. Symposium on Machine Learning and the Law at the29th Conference on Neural Information Processing Systems (NIPS 2016). Barcelona, Spain.

Hacker, P. (2018). Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic discrimination under EU law. Common Market Law Review 55(4).

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. Minds and Machines 30(1), 99–120.

Hamington, M. and Sander-Staudt, M. (2011). Applying care ethics to business. Springer, Oxford.

Hamington, M. (2019). Integrating care ethics and design thinking. Journal of Business Ethics 155, 91-103.

Harbison, J. (1992). Gilligan: a voice for nursing? Journal of Medical Ethics 18(4), 202–205.

Hasan, M. Macdonald, G. Ooi, H. H. (2022). How Facebook Fuels Religious Violence. Foreign Policy: https://foreignpolicy.com/2022/02/04/facebook-tech-moderation-violence-bangladesh-religion/ Accessed February 13 2022.

Held, V. (2006). The ethics of care: Personal, political, and global. Oxford University Press on Demand.

Henderson, G.R. and Williams, J.D. (2013). From exclusion to inclusion: An introduction to the special issue on marketplace diversity and inclusion. Journal of Public Policy & Marketing 32, 1_suppl, 1-5.

Hersh, M. A. (2016). Engineers and the other: the role of narrative ethics. AI & Society 31(3),327-345

Hoffmann, A. L. (2019). Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse. Information, Communication & Society 22(7), 900–915.

Iliadis, A. and Russo, F. (2016). Critical data studies: An introduction. Big Data & Society 3(2).

Jobin, A., Ienca, M., Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence 1(9), 389–399.

Johnson, G. M. (2022). Excerpt from are algorithms value-free? Feminist theoretical virtues in machine learning. Forthcoming in Ethics of Data and Analytics, Taylor & Francis.

Johnson, D. G. (2006). Computer systems: Moral entities but not moral agents. Ethics and Information Technology 8(4), 195–204.

Johnson, D. G. (2015). Technology with no human responsibility? Journal of Business Ethics 127(4), 707–715.

Johnson, D. G. and Powers, T. M. (2005). Computer systems and responsibility: A normative look at technological complexity. Ethics and Information Technology 7(2), 99–107.

Keats Citron, D. (2018). Sexual privacy. Yale LJ 128. HeinOnline, 1870.

Kelley, S. (2022). Employee perceptions of the effective adoption of AI principles, Journal of Business Ethics.

Kerr, I. and Earle, J. (2013). Prediction, preemption, presumption: How big data threatens big picture privacy. Stan. L. Rev. Online 66. HeinOnline, 65.

Kim, T. W., Maimone, F., Pattit, K., Sison, A. J., Teehankee, B. (2021). Master and Slave: the Dialectic of Human-Artificial Intelligence Engagement. Humanistic Management Journal 6(3), 355–371.

Kim, T. W., and Mejia, S. (2019). From artificial intelligence to artificial wisdom: what Socrates teaches us. Computer 52(10), 70–74.

Kim, T. W. (2018). Gamification of labor and the charge of exploitation. Journal of Business Ethics 152, 27-39.

Koehn, D. (1995). A role for virtue ethics in the analysis of business practice. Business Ethics Quarterly 5(3), 533–539.

Koehn, D. (1998). Virtue ethics, the firm, and moral psychology. Business Ethics Quarterly 8, 497-513.

Koehn, D. (2011). Care ethics and unintended consequences In M. Sander-Staudt and M. Hamington (eds). Applying care ethics to business, (pp. 141-153). Springer, Oxford.

Kohlberg, L. (1981). Essays on moral development. Harper & Row, New York.

Lepri, B., Oliver, N., Letouzé, E., Pentland, A., Vinck, P. (2018). Fair, transparent, and accountable algorithmic decision-making processes. Philosophy & Technology 31(4), 611–627.

Lin, Y.-T., Hung, T.-W., Huang, L. T.-L. (2021). Engineering equity: How AI can help reduce the harm of implicit bias. Philosophy & Technology 34(1), 65–90.

Lum, K. (2017). Limitations of mitigating judicial bias with machine learning. Nature Human Behaviour 1(7), 1.

Martin, K. E. and Freeman, R. E. (2004). The separation of technology and ethics in business ethics. Journal of Business Ethics 53(4), 353-364.

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

Martin, K. (2019b). Designing ethical algorithms. MIS Quarterly Executive 18(2), 129-142.

Martin, K. (2022a). Algorithmic Bias and Corporate Responsibility: How companies hide behind the false veil of the technological imperative. Forthcoming in Ethics of Data and Analytics, Taylor & Francis.

Martin, K. (2022b). Creating Accuracy and The Ethics of Predictive Analytics. http://dx.doi.org/10.2139/ssrn.3962551.

Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. Ethics and Information Technology 6(3), 175–183.

Melé, D. (2014). "Human quality treatment": Five organizational levels. Journal of Business Ethics 120(4), 457-471.

Mitchell, R.K., Agle, B.R., Wood, D.J. (1997). Toward a theory of stakeholder identification and salience: defining the principle of who and what really counts. Academy of Management Review 22, 853-886.

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

Mohammad, S.M. (2021). Ethics Sheets for AI Tasks. arXiv preprint arXiv:2107.01183.

Neubert, M. J. and Montañez, G. D. (2020). Virtue as a framework for the design and use of artificial intelligence. Business Horizons 63(2), 195–204.

Noddings, N. (1984). Caring. A feminine approach to ethics and moral education. University of California Press, Berkeley, CA.

Noddings, N. (2013). Caring. A relational approach to ethics and moral education. University of California Press, Berkeley, CA.

O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Broadway Books.

Onuoha, M. (2018). Notes on algorithmic violence. Retrieved from https://github.com/MimiOnuoha/On-Algorithmic-Violence. Accessed February 2022

Phillips, R. A. and Reichart, J. (2000). The environment as a stakeholder? A fairness-based approach. Journal of business ethics 23(2), 185-197.

Poole, S., Grier S., Thomas, K., Sobande, F., Ekpo, A., Trujillo, L., Addington, L., Weekes-Laidlow, M., Henderson, G. (2021). Operationalizing critical race theory (CRT) in the marketplace. Journal of Public Policy and Marketing 40(2), 126-142.

Rahwan, I. (2018). Society-in-the-loop: programming the algorithmic social contract. Ethics and Information Technology 20(1), 5–14.

Reader, S. (2007). Needs and moral necessity. Routledge, New York.

Rodgers, W., Nguyen, T. (2022). Advertising benefits from ethical artificial intelligence algorithmic purchase decision pathways. Journal of Business Ethics.

Ronald, K M., Agle, B. R., and Wood, D. J. (1997). Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts. Academy of management review 22(4), 853-886.

Rudner, R. (1953). The scientist qua scientist makes value judgments. Philosophy of Science 20(1), 1-6.

Sandin, P. (2009). Approaches to ethics for corporate crisis management. Journal of Business Ethics 87(1), 109–116.

Santoni de Sio, F., and van Wynsberghe, A. (2016). When should we use care robots? The nature-of-activities approach. Science and Engineering Ethics, 22(6), 1745–1760.

Segun, S. T. (2021). Critically engaging the ethics of AI for a global audience. Ethics and Information Technology, 23(2), 99–105.

Seele, P., Dierksmeier, C., Hofstetter, R., Schultz, M. D. (2021). Mapping the ethicality of algorithmic pricing: A review of dynamic and personalized pricing. Journal of Business Ethics, 170, 697-719.

Shaw, D., McMaster, R., Newholm, T. (2016). Care and commitment in ethical consumption: An exploration of the 'attitude–behaviour gap.' Journal of Business Ethics, 136(2), 251–265.

Shilton, K., Moss, E., Gilbert, S. A., Bietz, M. J., Fiesler, C., Metcalf, J., Vitak, J., and Zimmer, M. (2021) Excavating awareness and power in data science: A manifesto for trustworthy pervasive data research. Big Data & Society 8(2), 20539517211040759.

Simola, S. (2003). Ethics of justice and care in corporate crisis management. Journal of Business Ethics 46(4), 351–361.

Sison, A. J. G. and Redín, D. M. (2021). A neo-aristotelian perspective on the need for artificial moral agents. AI & Society. Published online: https://link.springer.com/article/10.1007/s00146-021-01283-0

Sison, A. J. G. (2015). Happiness and virtue ethics in business. The ultimate value proposition, Cambridge, UK: Cambridge University Press: Cambridge, UK.

Skinner-Thompson, S. (2020) Privacy at the Margins. Cambridge University Press.

Solomon, R. C. (1992). Corporate roles, personal virtues: An Aristotelean approach to business ethics. Business Ethics Quarterly 2(3), 317–339.

Soper, S. (2021). Fired by bot at amazon: 'It's you against the machine.' Bloomberg. https://www.bloomberg.com/news/features/2021-06-28/fired-by-bot-amazon-turns-to-machine-managers-and-workers-are-losing-out. Accessed 6 October 6 2021.

Stahl, B. C. (2021). AI Ecosystems for Human Flourishing: The Recommendations (pp. 91–115). Springer.

Steelman, T. A. and Rivera, J. (2006). Voluntary environmental programs in the United States: Whose interests are served? Organization & Environment 19(4), 05-526.

Steinberg, E. (2021). Run for your life: The ethics of behavioral tracking in insurance. Journal of Business Ethics.

Telkamp, K. B., Anderson, M. H. (2022). The implications of diverse human moral foundations for assessing the ethicality of artificial intelligence. Journal of Business Ethics.

Tronto, J. C. (1993). Moral boundaries: A political argument for an ethic of care. Routledge, NY.

Vallor, S. (2010). Social networking technology and the virtues. Ethics and Information Technology 12(2), 157–170.

Vallor, S. (2012). Flourishing on facebook: virtue friendship & new social media. Ethics and Information Technology 14(3), 185–199.

Vallor, S. (2015). Moral deskilling and upskilling in a new machine age: Reflections on the ambiguous future of character. Philosophy & Technology 28(1), 107–124.

Vallor, S. (2016). Technology and the virtues: A philosophical guide to a future worth wanting. Oxford University Press.

Vallor, S. (2017). AI and the Automation of Wisdom. In T. M. Powers, Philosophy and computing, (pp. 161–178). Springer.

Van de Poel, I., Nihlén Fahlquist, J., Doorn, N., Zwart, S., Royakkers, L. (2012). The problem of many hands: Climate change as an example. Science and Engineering Ethics 18(1), 49–67.

Van de Poel, I., Royakkers, L. M., Zwart, S. D., De Lima, T. (2015). Moral responsibility and the problem of many hands. Routledge, New York.

Van Wynsberghe, A. (2016). Service robots, care ethics, and design. Ethics and Information Technology,18(4), 311–321.

Vidgen, R., Hindle, G., Randolph, I. (2020). Exploring the ethical implications of business analytics with a business ethics canvas. European Journal of Operational Research 281(3), 491–501.

Villegas-Galaviz, C. and Martin, K. (2022). Moral distance, AI, and the ethics of care. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4003468

Villegas-Galaviz, C. (2022a). What the ethics of care is not. Retrieved from: https://carolinavillegasgalaviz.com/2022/05/02/what-the-ethics-of-care-is-not/ Accessed May 2022.

Villegas-Galaviz, C. (2022b). Ethics of Care as Moral Grounding for AI. In Martin, K. (ed.) Ethics of Data and Analytics. Taylor & Francis.

Waldman, A.E. (2021). Industry unbound: The inside story of privacy, data, and corporate Power. Cambridge University Press.

Waldman, A.E. (2019). Law, privacy, and online dating:“Revenge porn” in gay online communities. Law & Social Inquiry 44(4). Cambridge University Press, 987–1018.

Wicks, A.C., Gilbert, D.R. Jr., Freeman, R.E. (1994). A feminist reinterpretation of the stakeholder concept. Business Ethics Quarterly 4(4), 475-497.

Zhang, Y. and Zhou, L. (2019). Fairness Assessment for AI in Financial Industry. 33rd conference on neural information processing systems, Vancouver, Canada.

TABLE

Table 1 should appear after the section "**Deontological or Principle-based Ethics**"