FACULTY OF ECONOMICS AND BUSINESS ADMINISTRATION

(ICADE)

BUSINESS ETHICS AND ETHICS OF CARE IN ARTIFICIAL INTELLIGENCE

by

Carolina Villegas Galaviz

Advisors:

José Luis Fernández Fernández, Ph.D.

Kirsten Martin, Ph.D.

Madrid

April 2022

ACKNOWLEDGEMENTS

I was eight months pregnant with my second child when I started my Ph.D. Five months after the childbirth, the COVID-19 pandemic broke out in the world, bringing times of uncertainty and sadness for so many losses. Now, after two years of hard times, the invasion of Ukraine comes to (once again) remind us how dependent and vulnerable we are.

This dissertation is the product of years of reflection in challenging times. Times that make us reflect on how much we need ethics and care in this world, how we live in a web of interdependent relationships (such as Carol Gilligan explained), and how we can only become our better selves with the help and care of the other. Therefore, I want to dedicate my acknowledgments to giving thanks to the people without whom I would not be writing this or finishing my doctoral thesis.

I wouldn't be finishing this without the help of my advisors, Kirsten and José Luis. Thank you for saying yes to joining this project and believing in it not as a requirement but as something worth it. Thank you for teaching me the way to make it happen. Also, I want to thank both of you for everything you did that I couldn't see because you were so generous that you never complained about (fighting for maternity leave, founding, visas, agreements, and so long). Thank you for all your care.

I wouldn't be in this moment without the help of the academic world and the feedback from conferences, calls with scholars, workshops, and reviewers. Nor without the help of the administration of the Comillas Pontifical University and ICADE. I'll be forever grateful for the support from the Iberdrola Chair in Economics and Business Ethics, and from the Notre Dame – Technology Ethics Center at the University of Notre Dame. Also, I thank the education that I received during my years as an undergraduate and graduate student at the University of Navarra.

I wouldn't be here without my parents' care, love, and encouragement. [Ni sin su valioso trabajo (dentro y fuera del hogar) que puso los cimientos para que hoy pueda estar escribiendo esto]. Also, I wouldn't be here without the inspiration and feedback of my two brothers, my sister, and my friends.

I play in a team, and I wouldn’t be finishing this without them:

I wouldn’t be finishing this without the support of Nico. Thank you for always taking care of me, in so many ways. Finally, I would never have finished this without my greatest source of inspiration, our kids. I hope to be contributing with this work, even if it is a minimum, to a more caring world for both of you. Thank you for all the playtime that I borrowed to finish this. It wasn’t that much, since as a Ph.D. mom once told me: some of the better ideas come in the playground, when you are playing with your kids. Not in the quiet of a library cubicle.

Thanks once again to all, this dissertation has a piece of each one of you.

«DIEU d'Abraham, DIEU d'Isaac, DIEU de Jacob»

non des philosophes et des savants.

Blaise Pascal

ABSTRACT

With the introduction of Artificial Intelligence (AI) to business, face-to-face interactions are minimized, and decisions are part of an opaquer process, now reduced to data, that firms do not always understand. Within this process, moral implications appear when those who develop and deploy AI ignore circumstances, vulnerabilities, and the specific harm that can be done to individuals. In this scenario, the ethics of care is an adequate framework to approach AI ethical problems since it brings to the forefront of ethical decision-making the relevance of context and the consideration of vulnerabilities, interdependence, and the *voice* or what the *other* has to say.

This dissertation folds a compendium of studies that together contribute to the intersection of three main areas: business ethics, AI ethics, and the ethics of care. This doctoral thesis is made up of three journal articles (one under review, and two in a *Revise & Resubmit* status) and three chapters in edited collections, already accepted or published.

Chapter 1 examines the prominent moral approaches to the ethics of AI, identifies the strengths and limitations of each approach to the field, and proposes normative approaches (there the ethics of care) focused on power and vulnerable stakeholders as needed within the examination of AI within business ethics. Chapter 2 offers a unique conceptual contribution with the identification and analysis of moral distance as a problem within AI, and the proposition of the ethics of care to mitigate the issue.

In this compendium I develop some categories of the ethics of care: interdependent relationships, context and circumstances, vulnerabilities, and *voice* (what the other has to say). There I propose these can help ameliorate some of the critical problems within AI ethics. I refer to these categories in three chapters (Chapter 1, Chapter 2, Chapter 6). The ethics of care is a contextualized moral theory that argues for the most marginalized. That is why some scholars have misunderstood its purpose with altruism, feelings of pity, or

partiality. I oppose this view and analyze some of the regular incorrect claims about this ethical theory in Chapter 3, where I refer to what the ethics of care *is not*.

The last part of this thesis entails three brief book chapters in edited collections from the field of business ethics. Chapter 4 is based on a literature review and focuses on previous literature regarding the ethics of care and stakeholder theory. Chapter 5 focuses on AI and corporate responsibility and presents a summary of *how and why firms are responsible for AI*. Lastly, Chapter 6 presents the problem of profiling and classification of people in AI and proposes again the ethics of care as moral grounding for AI to analyze the harm and impact of algorithmically rating and scoring people.

The underlaying objective of this compendium is to understand, from an ethical perspective, *what we lose, and what we put at stake when we delegate our decision processes to algorithms.* This comprehensive question, led to specific interrogations and ethical issues analyze across the different chapters of this dissertation. Therefore, I propose the ethics of care as a unique and much needed approach to mitigate some ethical problems of the specific attributes of how AI works: blocking empathy in those who decide (caused by the distance between development and deployment of AI and its impact), reinforcing systems of power, unfairly rating and profiling individuals with data, and marginalizing vulnerable stakeholders.

KEYWORDS: *Business Ethics, Artificial Intelligence Ethics, Ethics of Care*.

RESUMEN

Con la introducción de la Inteligencia Artificial (IA) en los negocios, las interacciones cara a cara se minimizan y las decisiones son parte de un proceso más opaco, ahora reducido a datos, que las empresas no siempre comprenden. Dentro de este proceso, las implicaciones morales aparecen cuando quienes desarrollan e implementan la IA ignoran las circunstancias, las vulnerabilidades y el daño específico que se puede causar a las personas. En este escenario, la ética del cuidado es un marco adecuado para abordar los problemas éticos de la IA, ya que pone en primer plano de la toma de decisiones éticas la relevancia del contexto y la consideración de las vulnerabilidades, la interdependencia y la *voz* o lo que el otro tiene que decir.

Esta tesis doctoral presenta un compendio de estudios que en conjunto contribuyen a la intersección de tres áreas principales: la ética empresarial, la ética de la IA y la ética del cuidado. La tesis está compuesta por tres artículos de revista (uno en revisión y dos en estado *Revise & Resubmit*) y tres capítulos en colecciones editadas, ya aceptados y en proceso de publicación.

El Capítulo 1 examina los marcos conceptuales más conocidos dentro de la ética de la IA, identifica las fortalezas y limitaciones de cada enfoque en el campo y finalmente propone enfoques normativos (incluyendo la ética del cuidado) centrados en los problemas de poder y los *stakeholders* vulnerables. El Capítulo 2 ofrece una contribución conceptual única con la identificación y el análisis de la "distancia moral" como un problema dentro de la IA, y la propuesta de la ética del cuidado para mitigar el problema.

En este compendio, se desarrollan algunas categorías de la ética del cuidado: relaciones de interdependencia; contexto y circunstancias; vulnerabilidades; y *voz* (lo que el otro tiene que decir). Además, se propone que estos pueden ayudar a mejorar algunos de los problemas críticos dentro de la ética de la IA. Se hace referencia a estas categorías

en tres capítulos (1, 2 y 6). La ética del cuidado es una teoría moral contextualizada que defiende a los más marginados. Es por eso por lo que algunos académicos han malinterpretado su propósito con altruismo, sentimientos de lástima o parcialidad. Esta investigación se opone a este punto de vista y analiza algunas de las afirmaciones incorrectas habituales sobre esta teoría ética en el Capítulo 3, donde se hace referencia a lo que la ética del cuidado *no es*.

La última parte de esta tesis consta de tres breves capítulos de libros en colecciones editadas del campo de la ética empresarial. El capítulo 4 está basado en una revisión de la literatura y se enfoca en aplicaciones de la ética del cuidado a la teoría de los stakeholderss. El Capítulo 5 se centra en la IA y la responsabilidad corporativa y presenta un resumen de cómo y por qué las empresas son responsables de la IA. Por último, el Capítulo 6 presenta el problema de la elaboración de perfiles y la clasificación de las personas en IA y propone nuevamente la ética del cuidado como base moral para que la IA analice el daño y el impacto de calificar y puntuar algorítmicamente a las personas.

El objetivo subyacente de este compendio es c*omprender lo que perdemos y lo que ponemos en juego cuando delegamos nuestros procesos de decisión a los algoritmos*. En ese contexto, la tesis propone la ética del cuidado como un enfoque único y muy necesario para mitigar algunos problemas éticos de los atributos específicos del funcionamiento de la IA: como el bloqueo de la empatía en quienes deciden (causada por la distancia entre el desarrollo y despliegue de la IA y su impacto) y el refuerzo de los sistemas de poder que afecta a los *stakeholders* más vulnerables.

PALABRAS CLAVE: *Ética Empresarial, Ética de la Inteligencia Artificial, Ética del Cuidado*.

TABLE OF CONTENTS

[1] Under review, *Journal of Business Ethics*.
[2] Revise & Resubmit, *AI & Society.*

[3] Revise & Resubmit, *Anuario Filosófico*

[4] In G. Faldetta; Mollona, E.; Pellegrini, M. (Eds.) (2022). *Philosophy for Business Ethics*. Palgrave Macmillan.
[5] In Poff, D. and Michalos C. M. (Eds.) (2022). *Encyclopedia of Business and Professional Ethics*. Springer Nature.
[6] In Martin, K. (Ed.) (2022). *The Ethics of Data and Analytics*. Taylor & Francis.

INTRODUCTION

1. RESEARCH TOPIC

Data and analytics are the leading forces of technology shaping business and society. The process of business (and of any organization) decision-making is now completely overtaken by the hegemony of data. Data analytics continues to be the source where firms and organizations search for responses and proposals. Hence, questions about the ethical implication of data and Artificial Intelligence (AI) are critical to the good development of this technology and how it impacts business and society.

The broad spectrum of AI ethics as a field now conforms many different perspectives and approaches from various disciplines, including but not limited to philosophy (Anderson and Anderson, 2011; see also Liao, 2020), bioethics (Ekmekci and Arda, 2020), law (Barocas & Selbst, 2016), education (Prisloo, 2020; Slade and Prinsloo 2013, 2017), and business (Martin, 2019a, 2019b). This field has focused on problems such as privacy (Lane et al. 2014), bias and discrimination (Friedman and Nissenbaum, 1996; Barocas & Selbst, 2016), explicability (Floridi et al. 2018), transparency (Turilli and Floridi, 2009), or the search for principles and AI guidelines (Mittlestadt, 2019).

However, while many other fields study the ethics of AI, data, and analytics, business ethics is in a unique position to both normatively study AI and the responsibility of business. Scholars in the field of AI ethics and business have focused on ethical issues regarding the firm decision, its impact, and moral implications. Studies have focused on pricing (Seele et al. 2021), social media addiction (Bhargava and Velasquez, 2021), gamification (Kim and Werbach, 2018), unemployment (Kim and Scheller-Wolf, 2019), the effectiveness of principles (Kelley, 2022), responsibility (Martin, 2019b; 2019a) and the like.

Building on the work of AI ethicists, specifically of scholars in the field of AI and business ethics, the underlying objective of this dissertation is to understand and respond

to the question of *what we lose and what we put at stake when we delegate our decision processes to algorithms?* (From an ethical perspective). This comprehensive question led to specific interrogations and normative concerns analyzed across the different chapters of this dissertation. My thesis is the proposition of the ethics of care as a unique and much-needed approach to mitigate some ethical problems of the specific attributes of how AI works: blocking empathy in those who decide (caused by the distance between development and deployment of AI and its impact), reinforcing systems of power, unfairly rating and profiling individuals with data, and marginalizing vulnerable stakeholders.

Throughout this dissertation, mentions to AI refer to how "algorithms sift through data sets to identify trends and make predictions." (Martin, 2019b). 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). Hence, in the essence of AI is the need to frame individuals and situations statistically. When there is a new individual or scenario, algorithms confer the frame of those individuals or situations statistically related and decide accordingly. This appears as the problem of predictive analytics (Martin, 2021).

AI processes have the potential to improperly disregard legally protected classes and lead to a *disparate impact* (Barocas & Selbst, 2016). Where disparate impact "refers to policies or practices that are facially neutral but have a disproportionately adverse impact on protected classes."

The research question of this thesis boils down to one of the issues identified by professors Friedman and Nissenbaum (1996): that of the problems that originate "from attempts to make human constructs such as discourse, judgments, or intuitions amenable

to computers: when we quantify the qualitative, discretize the continuous, or formalize the nonformal."

In general, this doctoral thesis has the objective to defend the relevance of the qualitative, the continuous, and the nonformal in AI decision-making, and to defend that overlooking those is a cause of the automation of harm. Hence, the chapters of this dissertation identify and analyze (in different ways) some of the ethical problems derived from disregarding those human constructs. Within this process, moral implications appear when those who develop and deploy AI ignore *circumstances*, *vulnerabilities*, and the *specific harm* that can be done to individuals. There, this dissertation aims to explain that with the introduction of AI to business, face-to-face interactions are minimized, and decisions are part of an opaquer process, now reduced to data, that firms do not always understand.

For example, the use of AI reduces interactions and prevents those who develop and deploy AI from being aware of the impact of their actions, blocking empathy, and creating the problem of moral distance. Another example is how not everyone can be well represented in a model, and if we ignore that fact, AI will punish those who do not fit within the pattern.

Different normative approaches have proposed solutions to some of the challenges that arise from the introduction of AI within business. Among these theories are deontology and a focus on finding the right guide to AI ethics and principles to practice and business (Mittelstadt, 2019; Jobin et al. 2019). Also, propositions have come from other normative approaches such as justice and fairness (Lepri, 2018; Binns, 2018), responsibility (Johnson, 2015; Johnson and Powers, 20015), and virtue ethics (see Vallor, 2016).

The contribution of this dissertation is to propose the ethics of care approach to mitigate some of the ethical problems of AI within business. Issues arise from the said logic of algorithmic decision-making which entails blocking empathy in those who decide, unfairly rating and profiling individuals, or marginalizing vulnerable stakeholders.

The ethics of care was named as a theory and notion with Carol Gilligan in her pioneering study *In a different voice* (1982). Soon after that, Nel Noddings (1984, 2013) published *Care: A feminine approach to ethics and moral education* (which changed *feminine* to *relational* in its title in 2013). "This conception of morality as concerned with the activity of care centers moral development around the understanding of responsibility and relationships, just as the conception of morality as fairness ties moral development to the understanding of right and rules." (Gilligan, 1982). This novel proposition regarding morality constructs moral decisions as something that should be done in the understanding that individuals live in a web of interdependent relationships, and ethics should not be approached as a matter of contest of rights and a fight for equality.

Four decades have passed since the term was coined, and scholars in the field have developed a broader understanding of an ethic of care. In this dissertation, appears continuously the definition of the ethics of care proposed by Daniel Engster, one of the prominent scholars in the field (Engster, 2011; see also Engster, 2007). He defined the ethics of care in the following way:

> "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."

Daniel Engster proposed this definition in the study of the ethics of care from the stakeholders theory perspective, in a context of business, and that is why it is adequate

for the purposes of this thesis. This work was published in the edited collection *Applying care ethics to business* (Hamington and Sander-Staudt (Eds.), 2011), which bolsters this thesis.

The ethics of care is a contextualized moral theory that advocates for the most marginalized. That is why some scholars and practitioners have misunderstood its propositions with altruism, feelings of pity, and partiality. In this dissertation, I oppose this view and analyze some of the regular incorrect claims about the ethics of care in Chapter 3.

In general, this dissertation is at the intersection of three main areas: Business ethics, AI ethics, and the ethics of care. Hence, this thesis contributes to these three main disciplines, with a focus on business. Constantly, the contribution comes as questions that should be asked for those who design, develop, and deploy AI to its use in firms. Thus, this work entails both theoretical and practical implications. Those questions are based on four categories of the ethics of care that are crucial to mitigate the proposed AI ethical problems: Interdependent relationships; context and circumstances; vulnerability; and *voice* (or what the *other* has to say). Three of the six chapters use these categories literally (Chapter 1, Chapter 2, and Chapter 6).

Throughout all this document, the contribution is illustrated with examples of the practice (like Amazon firing algorithm, Microsoft chatbot Tay, or algorithms of *learning analytics* and hiring) that serve to explain the particular AI ethical issue and its impact within firms, as well as the way to apply the ethics of care categories and its propositions to ameliorate the issue.

AI responds to a logic of profiling and classification and a seek (even utopic) of objectifying decision-making, and that process is disregarding everything for which the ethics of care advocate. Hence, to propose the ethics of care approach as a new focus to

AI ethics within business may seem an impossible task, but in that apparent incompatibility appears the cause and reason to undertake the project, a much-needed one.

2. OVERVIEW

Each chapter of this compendium is an independent unit; and each explore a different issue, but always related to the overall objective of this dissertation. As independents parts, in each chapter, one can find an introduction to AI, the general problem and context; an explanation of the ethics of care and what it can bring to AI ethics and business; and how the proposed categories can help ameliorate the presented issue.

The reader should be aware that similar explanations are found between the chapters (of AI, the ethics of care, the categories, and other relevant concepts and definitions). Also, references are listed in each chapter following the guidelines of the publication where they pertain. Figure 1 is an illustration of the intersection of disciplines where this dissertation contributes. Also, figure 2 is an illustration of the dissertation structure, the chapters, and the concepts presented in those.

Figure 1. Dissertation intersection

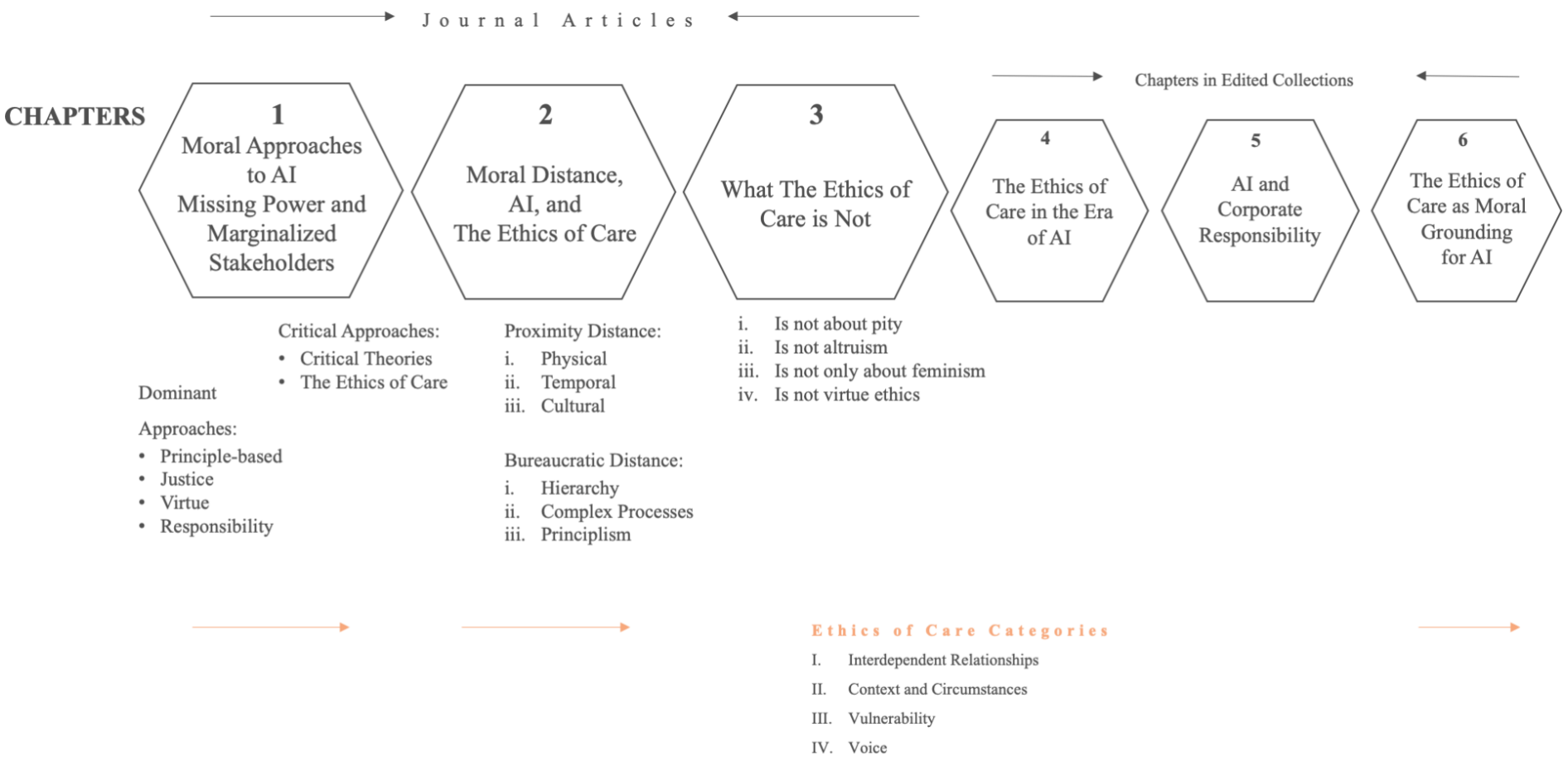

Figure 1. Dissertation Structure

Therefore, this dissertation is structured as follows:

2.1. Chapter 1

This chapter aims to analyze the prominent normative approaches to AI, identify the associated questions those approaches seek to address, and the limitations that each one may encounter. Here we examined four dominant normative approaches to AI: principle-based approach, justice, virtue ethics, and responsibility. We explain what each approach offers to the phenomena we analyze and their limitations. Here, limitations are not presented as *objections*, but as a way to express how theories are complementaries. There, we defend that unique attributes of AI– reinforcing systems of power, surreptitious, pervasive data collection, marginalizing vulnerable stakeholders– can be better addressed through specific normative approaches that raise the voice of the marginalized stakeholders either by focusing on the power dynamics of the more extensive socio-technical system or by prioritizing relationships between actors and their vulnerabilities. We propose to add critical approaches and the ethics of care, theories that offer a unique approach to critically examining AI within business ethics. These approaches center marginalized stakeholders, discussions around vulnerabilities and relationships, and the power dynamics of the current socio-technical systems for AI.

This chapter, written with Professor Kirsten Martin as a journal article, is under review in the *Journal of Business Ethics*. The article has been accepted for presentation in the conference of the International Society of Business, Economics, and Ethics, 2022; the 2022 Conference of the International Association of Business and Society (IABS); the 2022 Conference of the Institute Enterprise and Humanism (University of Navarra); and the Annual Meeting of the Society for Business Ethics 2022.

2.2. Chapter 2

This chapter investigates how the introduction of AI to decision-making increases moral distance and recommends the ethics of care to augment the ethical examination of AI decision-making and mitigate the issue. Within decision-making research, moral distance is used to explain why individuals behave unethically towards those who are not seen. Moral distance abstracts those who are impacted by the decision and leads to less ethical decisions. The purpose of this chapter is to identify the moral distance created by AI, caused by the distance between the development and deployment of AI and its impact. We conceptualize that the use of AI contributes to moral distancing in two ways. First, with the elimination of the face-to-face interactions (creating a distance of space, time, and culture), AI creates *proximity distance*. Second, the use of AI creates what we call *bureaucratic distance* derived from hierarchy, complex processes, and principlism. In order to help ameliorate the moral distancing from the use of AI, we propose the ethics of care and explain how four categories of the ethics of care (interdependent relationships, context and circumstances, vulnerability, and *voice*) are vital components to reduce the issues related to moral distance.

This chapter is a co-authorship with Professor Kirsten Martin and is currently in a Revise & Resubmit status in the journal *AI & Society*. The article has been presented in the following conferences and workshops: 2021-22 Zicklin Center Workshop in Normative Business Ethics of The Wharton School, University of Pennsylvania (October 2021); The Notre Dame-Carnegie Mellon University, Paper Development Workshop (of Technology and Business Ethics, October 2021); The *Society for Business Ethics Annual Meeting*, 2021; The International Vincentian Business Ethics Conference, 2021; The Notre Dame – Technology Ethics Center (ND-TEC) Affiliated Faculty Workshop Series

(October 2021); and the ND-TEC Undergraduate Affiliated Students Workshop Series (March 2022).

2.3. Chapter 3

To avoid some of the most common misunderstandings regarding the ethics of care, chapter 3 aims to identify what the ethics of care *is not*. For this purpose, I analyze four main misunderstandings about this moral theory: the ethics of care is not altruism, is not about partiality, is not only about women, and finally, is not a part of virtue ethics, but rather those are two different theories.

This paper is currently in a *Revise & Resubmit* status in the journal *Anuario Filosófico*. The objective of this work arises as a response to multiple misunderstandings that arose on various occasions whenever I presented my research project at conferences or workshops. I am grateful for the valuable feedback of professors Daniel Engster and Sheldene Simola, and for uncountable conversations with Nicole McAlee. Also, I am grateful for the feedback, comments, and edits of Maurice Hamington and my advisors.

The last part of this compendium entails three brief book chapters in edited collections from the field of business ethics.

2.4. Chapter 4

This chapter is the first approach in temporal terms to the research topic of this compendium. The chapter is based on a systematic literature review that examined all the references indexed in the Web of Science until January 2020. There, we collected the data using the keywords “artificial intelligence” and “ethics”. We found 1,370 documents, and finally ended up selecting 262 study units from journals in the three main disciplines of

the study: management, philosophy (ethics), and technology. This study approaches stakeholders' role in the AI ethics field and finally proposes a principle (based on the ethics of care and stakeholders' theory) to firms using AI.

Professor José-Luis Fernández-Fernández is the second author of this work, and a first version of it is forthcoming in *Philosophy for Business Ethics*. G. Faldetta, Mollona, E., Pellegrini, M. (Eds.) Palgrave Macmillan. Some of the content of this chapter was presented at the *Society for Business Ethics Annual Meeting*. 2020; The *British Academy of Management Conference*. 2020; the research seminars of ICADE in the Comillas Pontifical University, 2021; and the 2021 Symposium of the *Iberdrola Economics and Business Ethics Chair*.

2.5. Chapter 5

In chapter 5 our focus is on AI and corporate responsibility and summarizes *how and why firms are responsible for AI*. I second author this chapter with Professor Kirsten Martin and is forthcoming in the *Encyclopedia of Business and Professional Ethics*. Edited by Poff, D. and Michalos, C.M. in Springer Nature. The editors of this encyclopedia invited Professor Martin to collaborate on the book, and she invited me to join. We worked on this during my time as a visiting doctoral student at the Univesity of Notre Dame. Given the topic, the timing, and the relevance of the edited collection, we decided to incorporate this brief article as a chapter of my dissertation.

2.6. Chapter 6

The final chapter presents the problem of profiling and classifying people in AI and again proposes the ethics of care as moral grounding for AI to analyze the harm and impact of algorithmically rating and scoring people. After presenting the problem, I propose the four categories of the ethics of care that can help to ameliorate this issue. I

propose to apply this moral theory with some questions that those who develop and deploy AI should ask in their part of the processes.

This work is forthcoming in the book *Ethics of Data and Analytics*, edited by Professor Kirsten Martin with Taylor & Francis. Some of the content of this article was presented in the *European Business Ethics Network/Spain*, 2021; the *Conference of the International Association of Business and Society* (IABS), 2021; and in the class EG33999: Technology, Self, and Society at the University of Notre Dame in March, 2022.

The six chapters are the product of the feedback from all the conferences, and workshops in which those where presented. Also, the feedback of calls with professors that are experts in each of the three areas that bolster this dissertation. Furthermore, this dissertation has the input of the training and comments of doctoral workshops and symposiums of the Academy of Management (SIM Division); The British Academy of Management; the Society for Business Ethics, and the International Association for Business and Society. Also, it was especially illustrative the *Digital Footprint* seminar of the Fundación Pablo VI (which brought together academics, businesspeople, and government representatives) where one of the presenters talked about the problem of moral distance and AI; and the Care Ethics Lecture series where Carol Gilligan presented last January 2022.

REFERENCES

Anderson, M. and Anderson, S. (2011). *Machine Ethics*, Cambridge University Press, New York, NY.

Barocas, S., and Selbst, A.D. (2016). Big Data's Disparate Impact. California Law Review 104.

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.

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

Engster, D. 2011. Care ethics and stakeholder theory, in Hamington, M. & Sander-Staudt, M. (Eds.). *Applying care ethics to business*. New York: Springer: 93-110.

Engster, D. 2007. The heart of justice: Care ethics and political theory. Oxford: Oxford University Press.

Ekmekci, P. E. (2020). Artificial Intelligence and bioethics. Springer.

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

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Patrice, C., 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, 689-707.

Friedman, B. and Nissenbaum, H. (1996). Bias in Computer Systems, *ACM Transactions on Information systems*, 13(3).

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

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

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

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.

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

Kim, T.W. and Scheller-Wolf, A. (2019). Technological unemployment, meaning in life, purpose of business, and the future of stakeholders. Journal of Business Ethics 160, 319-337.

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

Lane, J., Stodden, V., Bender, S., Nissenbaum, H. (Eds.) (2014). Privacy, Big Data, and the Public Good. Cambridge University Press, Cambridge.

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.

Liao, S. M (Ed.). (2020). Ethics of Artificial Intelligence. Oxford University Press.

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

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.

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

Noddings, N. (1984). Caring: A feminine approach to moral education. University of California Press.

Noddings, N. (2013). Caring: A relational approach to moral education (2nd edition). University of California Press.

Prinsloo, P. (2020). Of 'black boxes' and algorithmic decision-making in (higher) education – A commentary. Big Data & Society January-June: 1-6

Prinsloo, P. and Slade, S. (2016). Student vulnerability, agency, and learning analytics: An exploration. Journal of Learning Analytics 3(1), 159-182.

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.

Slade, S. and Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. American Behavioral Scientist 57(10), 1510-1529.

Turilli, M. and Floridi, L. (2009). The ethics of information transparency. Ethics and Information Technology 11, 105-112.

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

CHAPTER 1

MORAL APPROACHES TO AI – MISSING POWER AND MARGINALIZED STAKEHOLDERS

Carolina Villegas Galaviz

Kirsten Martin

ABSTRACT

The introduction of AI to augment business decisions has strained the standard ethical approaches in business ethics, the firm is to focus on the interests of stakeholders. Unique attributes of AI and AI research – reinforcing systems of power, surreptitious yet pervasive data collection, and 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. The goal of this article is to examine the prominent moral approaches to the ethics of Artificial Intelligence (AI), identify the strengths and limitations of each approach to the field, and propose normative approaches focused on power and vulnerable stakeholders as needed within the examination of AI within business ethics.

Keywords: AI Ethics, Normative Approaches, Critical Theory, Ethics of Care

1.1 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 AI 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). Artificial Intelligence (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 of 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; Steinberg 2020), behavioral tracking (Steinberg 2017), social media addiction (Bhargava and Velasquez, 2021), 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. These limitations are presented not as objections (or replies), but as a way

to illustrate the unique contribution that each theory brings. Unique attributes of AI and AI research – reinforcing systems of power, surreptitious, pervasive data collection (Shilton et al. 2021), marginalizing vulnerable stakeholders– 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 stakeholders, 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.

1.2 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 value-laden decisions

of programmers and has moral implications for many when in use. In other words, decisions using AI models can diminish the rights of individuals, harm stakeholders, violate rules, norms, and laws, as well as unjustly distribute social goods (Martin, 2019). 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.

1.2.1. 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 guides 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). Some 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 am fulfilling the duty for transparency and beneficence?

While principles are helpful for ethics (Schwarz, 2005), they 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). Also, 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, following a principle-based approach, 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? Those who develop and deploy AI should ask, does this model represent my virtues and the character of a virtuous person? "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, partiality, or something only referred women.

1.2.2. 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 more fair processes. Algorithmic decision-making appeared more fair than "those made by humans who may be influenced by greed, prejudice, fatigue, or hunger" or any other feeling (Lepri et al. 2018). However, the AI ethics scholarship has succeeded to identify algorithms as a value-laden technology, that entails problems of bias, unfair representation, unresponsible accountability of mistakes, and the like (Martin, 2019). 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 & 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 & 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 systematic, 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 such as law (Hacker, 2017) or 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?

While fairness and justice approaches have shed much light on AI ethical issues, those also have 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 justice and fairness approaches, 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

1.2.3. Virtue Ethics

Virtue ethics is one of the main approaches in business ethics (Solomon, 1992; Koehn, 1995; Sison, 2014; Alzola, 2018). Primarily based on the Aristotelian propositions in the *Nicomachean Ethics*, this theory is usually related to the agent's character traits. Within business ethics, "a virtue ethical theory of business must be not only a normative theory about abstract principles and side constraints, but also a theory of practice that is accessible to the people whom business ethics is not just a subject of study but a way of life." (Alzola, 2018), while utilitarianism focusses on outcomes and deontological ethics on the action (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 about technology, specifically AI (Vallor, 2010; 2012; 2015; 2016; 2017; with Wallach, 2020). 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).

According to Vallor, 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), technologies such as AI. The answer is 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 a context impacted by technologies (Stahl et al., 2021; Kim and Mejia, 2019; Clark and Gevorkyan, 2020; Stahl, 2021). Also, there are some proposals to the inclusion of virtues in the design of AI models (Neubert, 2020; Wallach and Vallor, 2020; Gamez et al. 2020), of the virtue ethics approach as a framework for artificial moral agents (Sison and Redin, 2021), and as a critical factor to humans as master 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 Redin, 2021).

According to the virtue ethics scholarship, people within AI should question: 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, a virtue 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. As with the other approaches, this is not an objection to the theory (and does not imply that the theory disdains principles), but a limitation in the way that other theories may complement this approach with their focus, in this case, a principle-based approach with its focus on specific guidelines. 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 virtue ethics 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, does this process help or damage the character of the drivers? 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.

1.2.4. Responsibility

Normative approaches focused on responsibility issues examine the accountability of AI models and their impact. Within this approach, scholars focus on 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. 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 and Zwart 2015).

For AI, this problem entails the difficulty to hold accountability for the outcomes of a model between designers, developers, 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 Artificial Moral Agents (AMAs) and to differentiate voluntary actions from machine operations (Sison and Redin, 2021). Here to shed light on this complexity of attributing responsibility, appeared the notion of *Technological Moral Action*, which combined 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, according to 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 (Mathias, 2004), 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 would be a choice, not a result of technology's nature (Johnson, 2015; see also Sison and Redin, 2021). “Speculations about a responsibility gap misrepresent the situation and are based on false assumptions about technological development and about responsibility” (Johnson, 2015).

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.[2] Only by acknowledging the value-laden biases of algorithms can we begin to ask how companies inscribed those biases during design and development (Martin 2022). 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. While 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, 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.

2 Biases are value-laden design features with moral implications in use.

1.3. NORMATIVE THEORIES ABOUT POWER and the VULERNABLE

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, 2019b: 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., 2020). 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, 2019b).

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, Martin, and Parmar, 2020). Instead, subjects of the AI program are legitimate but marginalized stakeholders by being not only the most impacted, but also the stakeholders 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 stakeholders – 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.

1.3.1. 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., 2020; 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., 2020). A critical examination questions 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 this approach which problems of power balance appear when a bot automatically terminates drivers? And is the algorithm marginalizing the drivers? Or directly terminating marginalized groups? This approach also needs other focuses in the way to an ethical process in the firm.

1.3.2. 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 Lawrence Kohlberg, about six stages of moral development, women were not considered. Kohlberg (1977), 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). In her study, Gilligan discovered a *different voice*, and this voice has been part of women's socialization, which is why the ethics of care is related to feminism. However, Gilligan did not make essentialist claims about men and women. A *different voice* refers to a different way of moral deliberation which also extends to a 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 term was coined. 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, 2020; 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 stakeholders 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 design and engineering to create awareness of ethical decision making and the understanding of the 'other,' and within engineering 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). Withing technoscience, the ethics of care serve as a critical approach to emphasize responsiveness and ads 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), consumption (Shaw et al. 2016), or creative attitudes towards business (Alascovska and Bissonnette, 2019). Also, this approach has been proposed as a moral framework for stakeholder 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 stakeholders 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 (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 right 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 stakeholders?
- 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), partiality, or something limited to women (Villegas-Galaviz, 2022a).

In the example of Amazon, from an ethics of care and in the understanding of morality in a network of relationships, one should ask, how the terminations impact other members or groups of society? Also, 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? And what are the needs, issues, and interests of drivers?

1.4. 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 stakeholders – 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?

1.4.1. 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 widen the moral lenses used to examine AI to better foreground the marginalized and vulnerable stakeholders 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 stakeholders of the AI technology. For stakeholder theory, greater attention should be spent on those legitimate, urgent stakeholders with little power seen as discretionary or merely dependent stakeholders by Mitchell, Agle, Wood (Mitchell, Agle, Woods, 1997). As rightly noted,

firms will too frequently ignore such stakeholders while these are legitimate stakeholders with real concerns and interests. In the area of AI, these are also the stakeholders most impacted by the design and implementation of AI. Better named marginalized stakeholders, 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 stakeholders and how to bring their concerns into the design and implementation of AI by firms.

1.4.2. 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 stakeholders –. 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.

1.4.3. 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.

REFERENCES

Alacovska, A., and Bissonnette, J. (2021). Care-ful work: An ethics of care approach to contingent labour in the creative industries. Journal of Business Ethics, 169, 135-151.

Alzola, M. (2018). Character-based business ethics. In N.E. Snow (Ed.) The oxford handbook of virtue ethics. Oxford University Press.

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

Alacovska, A., and Bissonnette, J. (2021). Care-ful work: An ethics of care approach to contingent labour in the creative industries. Journal of Business Ethics, 169(1), 135–151.

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.

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., 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.

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.

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.

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.

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

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., & 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.

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.

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

Santoni de Sio, F., & 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.

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.

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

Waldman AE (2019a) 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.

CHAPTER 2

MORAL DISTANCE, ARTIFICIAL INTELLIGENCE, AND THE ETHICS CARE

Carolina Villegas Galaviz

Kirsten Martin

ABSTRACT

This chapter investigates how the introduction of AI to decision making increases moral distance and recommends the ethics of care to augment the ethical examination of AI decision making. With AI decision-making, face-to-face interactions are minimized, and decisions are part of a more opaque process that humans do not always understand. Within decision-making research, the concept of moral distance is used to explain why individuals behave unethically towards those who are not seen. Moral distance abstracts those who are impacted by the decision and leads to less ethical decisions. The goal of this chapter is to identify and analyze the moral distance created by AI through both proximity distance (in space, time, and culture) and bureaucratic distance (derived from hierarchy, complex processes, and principlism). We then propose the ethics of care as a moral framework to analyze the moral implications of AI. The ethics of care brings to the forefront circumstances and context, interdependence, and vulnerability in analyzing algorithmic decision-making.

Keywords: Artificial Intelligence, AI, Moral Distance, Ethics of Care, AI Ethics

2.1 INTRODUCTION

> At close range the resistance to killing an opponent is tremendous. When one looks an opponent in the eye, and knows that he is young or old, scared, or angry, it is not possible to deny that the individual about to be killed is much like oneself. (Grossman 1995)

When talking about remote fighting and drones, the issue of moral distancing means the disappearance of the vulnerable face of the opponent, which apparently, makes it easier to kill (Coeckelbergh 2013; see also Cummings 2004). The distance created by technology blocks the empathy that would arise when seeing the face of the *opponent*. However, it does not exempt one from moral responsibilities. When deciding to give or not a loan or a mortgage, or to deny or grant an insurance premium, it could be easier to deny specific opportunities if, thanks to the use of emerging technologies, such as artificial intelligence (AI)[7], we do not see the vulnerabilities and specific characteristics of people. Is easier to *kill*. For example, is Amazon firing delivery drivers easier when the assessment and task is fully automated with AI? If so, what framework could help firms better see the impacts of those morally relevant decisions in design and use?

If the use of AI impacts moral distancing – where decision is reduced to data, ignores circumstances, vulnerabilities, and the specific harm that can be done to that individual – the firm would miss the moral implications of their decisions for which they are responsible. The firm is blinded to the impact of their decisions behind the veil of AI while also being responsible for those decisions. In fact, companies continue to repeat the same mistakes. While bias in facial recognition programs has been known for years (Buolamwini and Gebru 2018), Google and Facebook, for example, have struggled with

7 With AI we refer to algorithms that "sift through data sets to identify trends and make predictions" (Martin 2019b, p 836).

issues related to AI and race, when their facial recognition algorithms had labeled Black individuals as "primates" (Facebook in 2021). Companies apologized, but mistakes persist and may continue until dealing with them become a priority for their leaders (Mac 2021).

The purpose of this chapter is to identify the moral distance created by AI. We conceptualize that the use of AI contributes to moral distancing in two ways. First with the elimination of the face-to-face interactions (creating a distance of space, time, and culture), the use of AI creates *proximity distance*. Second the use of AI creates what we call *bureaucratic distance* derived from hierarchy, complex processes, and principlism. The quid is that "the very distance between an act and its ethical consequences (ethical distance) may also play a determining role – if not always in the same way – in the transition process" (Zyglidopoulos and Fleming 2008).

In order to help ameliorate the moral distancing from the use of AI, we propose the ethics of care (Gilligan 1982; and Noddings 1984) as a moral framework to analyze technology and AI's moral implications. The notion of care within ethics promotes reference to vulnerabilities, context, and empathy in ethical decision-making (Gilligan 1982; Noddings 1984; see also French and Weis 2000; Held 2006). Within technology ethics, ethics of care may be the way to foreground notions of culture, diversity, and the "other" (Hersh 2016).

We argue that the ethics of care addresses the issue of moral distancing since the theory "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). However, the ethics of care does not imply deference to partiality, or feelings of patiality or altruism, but rather to take ethical reasoning beyond the reduction to principles and the consideration of purely quantifiable variables, and to

consider different points of view (hear different voices), interdependent relationships, context and circumstances, and individual vulnerabilities. The ethics of care is proposed as a complementary and integrative proposal to augment existing work in AI Ethics.

The chapter structure is as follows. First, we conceptualize the problem of moral distance, and explain how AI and technology exacerbates the issue. In the second part, we present the ethics of care, and emphasize four categories of the notion of care, there we explain how each of them could help to address the problem of moral distance.

2.2 MORAL DISTANCE AND AI

In his pioneering study, *Modernity and the Holocaust*, Zygmunt Bauman argues that some "moral sleeping pills," such as bureaucracy, may blur ethical concerns for those who are far from us, creating the problem of moral distance. Indeed, in some circumstances, distance may favor impartiality, or limit the possibilities of favoritism or influences of power. However, the issue of distancing in morality appears when "the natural invisibility of causal connections in a complex system of interaction, and the `distancing' of the unsightly or morally repelling outcomes of action" arise to the point of "making invisible the very humanity of the victims" (Bauman 1989). When developing his argument, Bauman takes the Milgram experiment, where the psychologist studied how obedience to authority could turn the participant to perform acts against their conscience and do great harm to another person.

Ethical implications of moral distance have been the center of studies focused on business decisions (Huber and Munro 2014; Mellema 2003; Zyglidopoulos and Fleming 2008; see also Jones et al. 2005). In the context of complex situations, such as organizational corruption in cases such as Enron, Arthur Andersen, and WorldCom, the concept of moral distance explains how individuals performed unethical acts even in

cases when they claim to have principles and values against those kinds of acts (see Mellema 2003; also Zyglidopoulos and Fleming 2008). There, the problem of moral distance is related to the conceptions of limits, is about boundaries and how they "demarcate not only physical, political, and other space but the moral space of inclusion and exclusion determining the limit and extent of our moral concern" (Chatterjee 2003). In that sense moral distance differs from moral disengagement, because in the former there is an actual distance that limits the whole comprehension of the moral context, whereas the latter refers to the belief of people which convince themselves that they are causing no harm or acting wrong, because ethical principles does not apply to them (Bandura 2002).

Mark Coeckelbergh (2013) was one of the first to directly addresses the relationship between distance, morality, and technology. In his work, Coeckelbergh takes the practice of using drones in remote fighting to illustrate "the claim that new technological practices that aim to bridge physical distance create more moral distance and make it difficult for people to exercise moral responsibility" (p. 88). Previously, Bauman (1989) also talked about the role of technology in increasing and exacerbating the problem of distance and morality when he explains that the issue "becomes particularly acute in our modern, rationalized, industrial technologically proficient society because in such a society human action can be effective at a distance, and at a distance constantly growing with the progress of science, technology and bureaucracy."

Moral distance has two components. The first is proximity, in space, time, and culture, where individuals tend to behave ethically regarding those in close proximity to them. The second component is related to bureaucracy and hierarchy; when a person's act is a small part of an extensive process, a kind of moral diffusion of responsibility appears, as in the problem of 'many hands,' leading to moral distancing. Moral distance could be

created by the temporal, physical, and cultural distance of a person between his or her act and its consequences, and it can be the result of an organization's bureaucracy (Huber and Munro 2014; Zyglidopoulos and Fleming 2008). We extend the examination of technology implications on moral distance addressing how AI exacerbates the problem.

2.2.1. Proximity Distance

According to Bauman (1989), there is an inverse ratio of readiness to cruelty and proximity to victims, "it is difficult to harm a person we touch. It is somewhat easier to afflict pain upon a person we only see at a distance. It is still easier in the case of a person we only hear. It is quite easy to be cruel towards a person we neither see nor hear." Moreover, many ethicists as Aristotle (in *Rethoric*) or Hume (in *A treatise of human nature*) have talked about the problem of proximity in morality and how distance affects moral action (Chatterjee 2003). "Ethical traditions that base morality on human nature claim that distance over time and place matters morally because humans are by nature unsuited to show equal concern to distant people and events compared to those near in time and place" (Chatterjee 2003, p. 327). However, distance does not exempt from moral responsibilities.

Regarding proximity distance, there might be a physical, temporal, and cultural impact in distancing.

2.2.1.1.Physical distance

Regarding *physical distance* affecting morality, Bauman (1989) argues that moral inhibitions are tied to human physical proximity, hence moral inhibitions may not act at a distance. For example, Bauman says that "the increase in the physical and/or psychic distance between the act and its consequences …quashes the moral significance of the act

and thereby pre-empts all conflict between personal standard of moral decency and immorality of the social consequences of the act." This implies that the physical distance allows the abstraction of the acting subjects, annulling (for them) the moral meaning of the act, that may be the cause of them to act against their principles.

In the field of technology, the physical distance created by information technology is described by Coeckelbergh's (2013) explanation of moral distance in pilots' remote fighting with drones. In remote fighting, it is easier for soldiers to kill; in body-to-body fighting, soldiers comprehend the opponent as a similar person, as an equal. The close contact opens the possibility of feeling empathy, which impacts the soldiers' decisions to kill. The author presents moral distance as a moral-epistemological problem since those using the specific technology (due to the distance facilitated by the tool) do not fully know the possible outcomes of their action. Bolstered on the work of Heidegger and Levinas regarding distance, technology, and morality, Coeckelbergh states that technology creates a new world for those using it: "The technology and the distance it creates does not only produce a barrier between our empathic capacity and the opponent, it changes the very way we perceive that opponent" (p. 93). Within the same argument, scholars have defended that technology has ethical implications because it limits engagement and commitment (Borgmann 1984), and is a tool to eliminate personal vulnerabilities (Dreyfus 2008).

Within AI the idea that algorithms can increase anonymity and psychological distance has been identified as a relevant threat in the possible malicious use of AI (Brundage et al. 2018). According to Brundage el al. (2018), many tasks involve communicating face-to-face, and "by allowing such tasks to be automated, AI systems can allow the actors who would otherwise be performing the tasks to retain their anonymity and experience a greater degree of psychological distance from the people

they impact." This means that proximity distance creates an epistemological or psychological distance with moral implications. Also, Coeckelbergh (2015) states that AI and automation alienate individuals from material reality since "we now work in ways that no longer require intense and direct engagement with that material reality … it thereby creates a gap, a distance, between us and nature."

For example, Amazon algorithmically rates their drivers and automatically fires them by email. When a human is asked to assess the ability of a driver, they may have been a driver themselves, understand what the data means in the context of a given route, and would be required to interact with the individual before telling them they are fired. However, if AI is the boss who fires the employee, the program does not know that the firm is firing an Army veteran who claims to have done nothing wrong in his job (Soper 2021) or a mother affected by the economic crisis of the COVID-19 pandemic to whom before being fired was told (from the same Amazon app) that she was doing a "great" job, in a scale of Fantastic, Great, Fair or At Risk (Gilbert 2021). If the whole process, from rating to firing, is automated, the company is ill-equipped to address specific circumstances affecting the data.

2.2.1.2. Temporal distance

Zyglidopoulos and Fleming (2008) describe *temporal distance* as the type that "refers to how far into the future the consequences of one's acts are. The further ahead in time these consequences are, the easier it will be for individuals to discount the moral consequences of their act." With this concept, the authors explain how the short-termism of business influences temporal distance. With a focus on immediate results, people lose sight of the future consequences of their actions. Also, Stephen M. Gardiner (2003) talked about temporal moral distance in line with what the author calls *intergenerational* ethics,

or the obligations of one generation to future people. The said means that there is a problem of moral distance when a person's acts affect (even without being fully conscious) future generations.

Regarding temporal distance and technology, due to the impact that the design of AI models can have in the near future and in future generations, those who develop AI should leave open the possibility to change variables that can take on different meanings over time. Hence, they allow for fairness for future generations. Within the discussion regarding the ethics of technology and time, the philosopher Hans Jonas (see Jonas 1984) is one of the prominent voices. According to Jonas, technology enlarges the impact of human action, and with what humans do here and now, thinking of their own goals, massively influences the lives of millions of people in other places and the future. Furthermore, those affected have no voice or vote in this regard (Jonas 1984).

Within AI the problem of this type of distancing appears in the temporal distance between the development of AI models and their deployment. Also, it appears in how models can change with the introduction of new data.

An example of this subtype of distance is the AI Microsoft's chatter bot Tay, released on Twitter in 2016. Since the bot learned to reply based on interaction, some hours after its introduction in the platform, it began to use offensive language.

2.2.1.3. Cultural distance

Nicholas Rescher (2003) explains the moral significance of cultural distance when talking about judging with external moral standards. The idea refers to how morality and ethical standards should not be separated from the culture where they emerge. Principles and values are bolstered in a historical-cultural context and one should not judge remote people by standards of different societies and cultures. According to Rescher, that does

not entail “an indifferentist relativism, but rather the contextualism of a situationally determinate value system” (p. 477).

In what refers to technology and cultural distance, there is a codependence between culture and technology where one influences the other, within that context and in conversation with Bauman’s propositions, Nørskoy (2021) proposed the idea of “asethical cleanising”, what he explains as a risk of sanitation of culture by science and technology. What that means is that the elimination of culture (with its norms and impact) in technology and the reduction of social and morally significant interactions to technological optimization and performance would lead society mistakenly imbued by moral correctness enforced by the robotic environment (Nørskoy 2021). This scenario would lead to a problem of cultural moral distance.

AI enlarges this moral problem by automating ethical decision-making and the fact that the same model may be used for different cultures. These models, developed from culturally-myopic training data, encode the relative importance of rules and principles that will be used to make decisions in the future. If principles and values are rooted in society, to no fall into a problem of moral distance, developers should consider the moral specificities of each culture. Many cultural factors contribute to moral judgments such as religion, demographics (like population density or economics), the history of the own culture, and the like (Graham et al. 2016). Scholars have argued for the need to engage AI ethics with diverse cultures (Segun 2021).

The example of *learning analytics*, which refers to the measurement, collection, and analysis of student data in higher education (Slade and Prinsloo 2013, 2017), illustrates this type of distance. With the knowledge that algorithmic decision-making has the power to shape social life, scholars in the field of ethics and *learning analytics*, argue that there are tensions in research on using algorithms to decide things like who gets

accepted into institutions or who can access student funding (Prinsloo 2020). Universities have always used quantifiable variables, like GPA, but AI exacerbates the moral distance that the blindness attachment to those variables can create.

A university would design an acceptance program that must accommodate applicants from different cultures. However, grades have different meanings in different countries. In the case of Spain, only one in twenty students obtain the highest grade, so the scale conversion to a country with a different system (in which more than one student can obtain the highest grade) rate a Spanish candidate lower. Also, the relevance that different variables have on a student's performance varies according to culture. Some cultures value extracurricular activities, volunteering, or networking, and for some others, these may not be significant. If an algorithm automatically decides without considering student nationalities, international students will lose a place they merit. Moreover, if a model frames applicants from the same culture or country as inappropriate, students from disadvantaged neighborhoods or countries may not be the right fit for acceptance according to some models. However, to automate that decision could have a great impact in the future with the marginalization of those areas, or entire societies, leaving them with no possibility of prospering.

2.2.2. Bureaucratic distance

The second form of moral distance is bureaucratic distance. For Bauman (1989), bureaucracy functions as a moral sedative since it is programmed to seek the optimal solution and "to measure the optimum in such terms as would not distinguish between one human object and another, or between human and inhuman objects. What matters is the efficiency and lowering of costs of their processing" (p. 104). Hence, bureaucracy is about procedures of formal rationality, with hierarchies, and of complex processes where

one's actions are a small part of a bigger objective and should reduce to follow scripts. Those scripts abstract “the real consequences of the defects … into a set of depersonalized figures and formulae” (Zyglidopoulos and Fleming 2008). The key idea is that proximity distance and bureaucratic distance contribute to creating an "inhuman context" (Huber and Munro 2014) that depersonalized those individuals who will be affected by the consequences of the decision. Within the bureaucratic moral distance, here we explain three subtypes of distancing: hierarchy, complex process or the problem of ‘many hands,’ and principlism.

2.2.2.1. Hierarchy

Hierarchy increases moral distance because individuals tend to act against their principles when an authority demands (as in Milgram’s experiment). Bauman (1989) identified the problem of hierarchy and bureaucracy as one of the main drivers of moral distance. Bauman says that in a bureaucracy what matters is “how smartly and effectively the actor fulfills whatever he has been told to fulfill by his superiors, [which] in addition to giving orders and punishing for insubordination, they also pass moral judgements – the only moral judgements that count for the individual's self-appreciation” (Bauman 1989). Individuals tend to hand over their responsibility for their action to those who have ordered them to carry it out and limit themselves to doing their chores in the way that they have been instructed.

The problem of moral distance and hierarchy is prominent in business and AI ethics, where the distance between layers of organizations appears between developers and companies. In this case, the effect of authority on people who develop algorithms continues to have the moral distance implications defended by Milgram and Bauman (1989). The quid is that developers “work in an environment which constantly pressures

them to cut costs, increase profit and deliver higher quality … Managers might coerce ICT professionals to make unethical or at least disputable decisions to the so-called benefit of the company" (Van den Bergh and Deschoolmeester 2010). Even though developers might have an established code of conduct and AI ethical guidelines, they face pressure from managers to design models that prioritize company interests (Mittelstadt 2019). There, some have argued about the threats of letting industry write the rules for AI (and sponsor AI ethics research) and a type of "emissions" of high-tech industry as to how their profits are borne by society (Benkler 2019).

This subtype of moral distance also appears in human deference to AI decision-making. "Humans, some argue, will happily defer to the machine. Yet such blind deference is ill founded" (Zarsky 2016). There, awareness of algorithmic shortcomings becomes critical for those deploying AI. In this type of issue, "the distance de-skills us: we become dependent on the technology and we do no longer know how it works, what it does, and indeed what we are doing" (Coeckelbergh 2013).

For example, in the case of Amazon, where a data analytics program rates and fires drivers, there is no human intervention in the decision, even the dismissal notice is automatic, and there are also no prior notification protocols so that drivers can appeal the decision before their termination. The company completely defers to the decisions made by the data analytics program.

2.2.2.2. Complex Processes

The second implication of bureaucracy and distancing emerges from *complex processes*. The latter implies the act of several persons, which we relate to the so-called problem of 'many hands.' Dennis Thompson (1980) was one of the first authors to state this issue. According to him, since many officials contribute in several ways to

government decisions and policies, is hard to identify who is morally responsible for those policies' outcomes, creating the problem of 'many hands.' In situations that involve several people's performance, individuals tend to diminish the ethical implications of their acts. Hence, "loosely, this problem may be described as the problem of attributing or allocating individual responsibility in collective settings." (van de Poel and Zwart 2015). Also, this problem of 'many hands' has two varieties: backward-looking (when something went wrong and the responsible is unknown) and forward-looking (when there is a need for a collective action to accomplish something). There are many real examples of the two varieties as financial crisis, global warming, or poverty, and on a small scale could be the bankruptcy of a company or the lack of communication of processes in an organization (van de Poel et al. 2015).

Bauman (1989) talks about the relationship of this problem to moral distance when he explains that "with most of the socially significant actions mediated by a long chain of complex causal and functional dependencies, moral dilemmas recede from sight, while the occasions for more scrutiny and conscious moral choice become increasingly rare."

The distance created from complex situations and the problem of 'many hands' is a recognized problem in the field of ethics of technology (van de Poel et al. 2012; Coeckelbergh 2020). Helen Nissenbaum (1996) correlates the problem to technology and explains its implications for moral responsibility, she defends that computer systems are products of groups of individuals and organizations, and if a "system malfunctions and gives rise to harm, the task of assigning responsibility – the problem of identifying who is accountable – is exacerbated and obscured." Also, that product finally impact the life of an end user.

For AI this issue appears in the problem of inscrutability where "the degree to which an algorithm is inscrutable contributes to our ability to identify, judge, and correct mistakes in algorithmic decisions" (Martin 2019a). This type of algorithmic opacity, "an opacity that stems from the mismatch between mathematical optimization in high-dimensionality characteristic of machine learning and the demands of human-scale reasoning and styles of semantic interpretation" (Burrell 2016) is critical to the understanding of how AI creates moral distance. The said implies that, if the logic of an algorithm is incomprehensible for those who deploy it, an insurmountable moral distance problem would appear.

For example, in the case of either the university admittance program or the Amazon program to rate and fire drivers, moral distance is created, between the manager and the individual impacted by the decision, due to a lack of understanding of how the decisions are made by the program.

2.2.2.3. Principlism

Finally, the third contributor to bureaucratic moral distancing is principlism. We refer to principlism as the kind of moral distance that creates a blind attachment to guidelines and principles, which sometimes ended up in unfairness. Huber and Munro (2014) first conceptualized this issue as *ethical violence* which they defined as a type of moral distance and a blind attachment to guidelines and principles, on those situations in which, by adhering to ethics, people end up interposing a distance with the specific circumstance. At first studies of moral distance described the problem as something of bureaucracy and formal rationality. However, research developed to outline the problem as something that can appear "even in informal personal relationships … and may be even

implicit in the notion of ethics itself … where ethical principles can serve to justify condemnation and cruelty in the name of ethics" (Huber and Munro 2014).

In the field of AI, this type of moral distance appears in how AI guidelines, recommendations, or standards proposed for the field, may not be a solution in several scenarios. Although the use of ethical guidelines in all types of organizations has been questioned, the effectiveness of ethical codes in the field of technology has a special skepticism (Van den Bergh and Deschoolmeester 2010; see also Bia and Kalika 2007). Mittelstadt (2019) argues about how *principles alone cannot guarantee ethical AI*, and how a principle-based approach "may have limited impact on design and governance."

Taking these characteristics of AI, within the idea that principles cannot guarantee the ethics of AI, Mittelstadt defends that "going forward, AI ethics must become an ethics of AI business," to avoid the abstraction of principlism. There "if robotics systematically enters into our interpersonal sphere, the symmetric reciprocal meeting … where the other is a unique and mortal individual (Levinas 2017a, d) and elicits an ethical imperative, risks being subverted by the reduction to some quantitative computational measures which are digitally representable and match whatever criteria implemented in or feeding the program running the robot" (Nørskoy 2021). What we defend here is that the "codification of ethics," (top-down, bottom-up, or hybrid approaches (Allen et al. 2005)) in AI may end up in a principlism moral distance problem.

For example, in the case of learning analytics, algorithms may be designed to avoid unfairness and promote equality, and there to evaluate prospective students with the same variables. However, to apply the same ratio, a principle, to students with different circumstances could be unfair to those students who do not fit the rule formalized in the program. By relying on the algorithm, developed by individuals and/or trained on historical data, to provide principles of admittance for future students, the

university prioritizes those rules over those students who do not fit the mold of the majority who are well represented in the data and algorithm. Table 1 is a summary of the types of moral distance and of how AI exacerbates the problem.

TABLE 1. Outline of the types of moral distance and the explanation of how AI exacerbates the problem.

Moral Distance	**Subtypes and Definition**		**With AI**
Proximity Distance	Physical	No face-to-face interactions.	- Decision-making without seeing/interacting with those affected by it.
	Temporal	How own actions impact the future (and future generations).	- How models can change with the introduction of new data.
	Cultural	Using the same value to judge different cultures.	- Deploying the same model in different cultures. - Framing subjects from different cultures using the same algorithm.
Bureaucratic Distance	Hierarchy	The power over subordinates.	- Influenced power of companies over developers. - Deference of decisions to AI as an authority.
	Complex Processes	The problem of many hands and blurred responsibility.	- Who is responsible for AI decision-making consequences? - Algorithmic inscrutability.
	Principlism	Reduction of morality to principles or a blind attachment to moral guides.	- How principles alone have a limited impact on AI ethics.

2.3. THE ETHICS OF CARE AS A BRIDGE FOR MORAL DISTANCE IN AI

We propose the ethics of care as a moral framework for AI ethics and the problem of moral distance. Carol Gilligan's and Nel Noddings are the pioneering scholars within the ethics of care. The theory was in response to the orthodoxy of ethics of justice, since the theory is not bolstered on inviolable impartial principles but appeals to relationships and context (French and Weis 2000; see also Held 2006). Nevertheless, is essential to address this theory as complementary and integrative approach in relation with other ethical theories such as virtue ethics or ethics of responsibility, to name a few.

The ethics of care asks us to focus on the concrete situation and provide answers concerning circumstances and context (Gilligan 1982). Hence, the ethics of care is framed as a moral vision centered on the individual. While other ethical theories such as deontology, utilitarianism, or consequentialism respond to ethical principles and duties, care ethics focuses on fostering people's vulnerabilities and needs (Weltzien and Melé 2009). Therefore, care should be understood as a *practice and a work that must be done on a direct level* (Sander-Staudt and Hamington 2011), as a value and as an activity. In this sense, it can be argued that the ethics of care proposes solutions according to the interests of each party and not to previously established norms (Reiter 1996).

The ethics of care has been proposed in situations where the interests of the least advantaged stakeholders are not being considered. In other words, where the distance between those making the decisions and those impacted by the decisions is too great and the marginalized stakeholder's interest are not being seen or judged to be legitimate. The notion of care has also been proposed in the field of technology ethics. Marion Hersh (2016) suggested the ethics of care (along with narrative ethics and virtue ethics) for engineers to make conscious ethical decision making, and for the understanding of notions of culture, diversity, and the "other", and issues related to these. According to

Hersh, the understanding of those notions is "very important for engineers for a number of reasons, including 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."

In the same line, Maurice Hamington (2019) proposed the integration of care ethics and design thinking, which is a practice from engineering (latter adapted to business) to "enable innovation and problem solving through participatory processes." Since the essential idea of design thinking is to consider those who will use the design, "including their emotions and ambiguities," the author proposed care ethics as a framework to help in the understanding of the end user of products and services.

The ethics of care have been applied to different areas in the business world since the 1990s (Melé 2014). There are also studies that propose ethics of care as moral framework for stakeholder theory (Wicks et al. 1994; Burton and Dunn 1996; and Engster 2011). Burton and Dunn (1996) proposed care as a moral grounding to stakeholder theory and management decision-making, and stated a principle: "Care enough for the least advantaged stakeholders that they not be harmed; insofar as they are not harmed, privilege those stakeholders with whom you have a close relationship." In essence, for the notion of care, firms must avoid any possible form of harm and take moral sentiments that we all share when making decisions (Wicks et al. 1994).

According to Daryl Koehn (2011), care ethics is the necessary framework in the context of unforeseen and unintended consequences because of the theory's flexibility and reference to interdependence, empathy, sympathy, and trust, rather than rules.

Since its establishment, the notion of care has developed from an ambiguous concept to a more rigorous definition. Daniel Engster (2011) proposed a definition of care ethics 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." Hence, the ethics of care goes in line with avoiding harm and taking vulnerabilities, relationships, and context, as well as emotions and empathy as appropriate guides to ethical decision-making (Sander-Staudt and Hamington 2011).

Care ethics involves attention and empathic response, a commitment to attend to legitimate needs (Noddings 1984)[8]. However, the ethics of care is not altruism or feeling sorry for someone or making decisions to do someone a favor. Rather, "although we often think of care in terms of characteristics such as understanding, responsiveness and nurturance, the practice of care is not always and necessarily harmonious … [and] … may be imbued with conflict" (Simola, 2010; see also Simola, 2015).

Within the study of the ethics of care, there exists a shared understanding that ethical decision-making should consider *interdependent relationships*, *context and circumstances*, attend to particular *vulnerabilities*, and also should respond to different points of view, hear different *voices*. We emphasize four categories of the ethics of care that may help to ameliorate the problem of moral distance.

2.3.1 *Interdependent relationships*

In the ethics of care, individuals and interests are not isolated but rather have meaning in a web of interdependent relationships. Within the ethics of care, we are to not only maintain relationships but also be responsive to the needs of others. Rather than a focus on principles, the ethics of care appears as a method and a way to orientate towards the world (Sander-Staudt and Hamington 2011).

8 Also, "care" should be distinguished from "personal service," the former "involves meeting the needs of those who are unable to meet such needs themselves, the latter involves meeting needs for others who could meet such needs themselves." (Sander-Staudt and Hamington 2011; see also Noddings 1984).

The quid of this category is that responsibility and morality can only be understood in a network of relationships, where one puts aside the general standard and look to the concrete situation, and there "the generalized other" becomes "the particular other," a specific individual in a particular circumstance. Also, "the ideal of care is thus an activity of relationship, of seeing and responding to need, taking care of the world by sustaining the web of connection so that no one is left alone" (Gilligan 1982). According to Nel Noddings (1984) this language "concentrates on relationships, needs, care, response, and connection rather than principles, justice, rights, and hierarchy."

When applying ethics of care to AI, it would be essential that models do 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). Developers would be in charge of the understanding of interdependent relationships and how they can be impacted by algorithms. An example of the application of this category can be illustrated in the field of learning analytics, when universities decided which students retain in a program using AI. With algorithm decision-making universities consider quantifiable variables but ignore essential facts that could also add value to the student's profile. For instance, a student with lower grades may have *interdependent relationships* impacting his or her scores, such as family dependents (as infants, grandparents, or parents with illness), but show commitment and responsibility. Also, this may be the case when students lower their performance because of the dedication of their time to volunteering or leadership of student associations.

2.3.2 *Context and circumstances*

The ethics of care is a practice and something to be done on a direct level and may be understood as a "motive, ideal, virtue, and method" (Stander-Staudt and Hamington 2011). Moreover, "care theorist generally agree that care is a relational approach to morality that entails contextualized responsiveness to particular others in a manner that supersedes the mere delineation of rules or calculated consequences" (Hamington 2019).

Consideration of context and circumstance would address the problem of proximity (space, time, and culture) in moral distance. Those working with AI would seek to understand the direct consequences their actions have on others, bridging the impact of "complex process" on moral distance. Those who develop and deploy AI are responsible for ensuring that algorithms do not eliminate variables such as context (like space and time), circumstances, or historical-cultural background.

For example, within learning analytics, this category appears as relevant when context and circumstances may affect someones' GPA and that its acceptance in a program. Students' grades can be lower in disadvantaged neighborhoods or countries if teachers and school supplies are not optimal, but someone who comes out in unprofitable circumstances has a lot of courage, strength, and determination. Also, in the case of Amazon firing bot, former Amazon managers explain that "the largely automated system is insufficiently attuned to the real-world challenges drivers face every day" (Soper 2021).

2.3.3 *Vulnerability*

Within the ethics of care, understanding vulnerability is vital to understand the relevance of the needs and suffering of others and to act according to those who can be affected by a decision. Developing AI based on the ethics of care implies that algorithms do not prevent individuals from meeting their needs. When applying ethics of care to AI,

those who develop and deploy AI could certify that algorithms do not prevent the possibility of fostering the needs of protected classes, people at risk of social exclusion, or marginalized stakeholders. Also, those working with AI could guarantee that the data used does not imply exploiting the vulnerabilities of those affected by the algorithm and that vulnerabilities are not used as variables to prevent their future enhancement.

For example, admittance decisions would want to address the specific *vulnerabilities* of students, such as a student with an attention deficit disorder which affects the student capacities in some subjects but not in others.

2.3.4 *Voice*

When utilizing care ethics, one would identify and hear the range of voices impacted by the decision. According to Carol Gilligan (1982), "to have a voice is to be human. To have something to say is to be a person. But speaking depends on listening and being heard; it is an intensely relational act." For Gilligan, in *In A Different Voice*, within decision-making and morality is critical to give voice to every affected part in any situation. Communication is the way to resolve ethical conflicts because when communicating, one hears different voices, opinions, and points of view. Hence, the needs of all those who are affected by an action are considered.

In the case of moral distance and AI, this notion addresses the types of proximity distance, giving voice to each culture and those far in space and time. Although AI decision-making eliminates physical interaction, moral issues created by that distancing could be addressed by listening to the voices of those affected, while those who develop AI models understand it as a priority to consider the interests of all parties.

This category could ameliorate the problem of proximity distance if algorithms maintain open the possibility of hearing different voices and not silence any voice that

should have been part of the situation in which is applied. There might be various formulas to give voice to all stakeholders, especially the most marginalized, like the work of interdisciplinary and intercultural teams working to develop and deploy AI. In the try to give voice to every affected part, those working on AI are considering the implications of the passage of time, the interests of those who are not present (whom they cannot see or who may never have a physical approach), and the needs and values according to different cultures.

For example, a university can lower the voice of marginalized groups with the application of a model which does not represent their situation or case very well. Similarly, the implementation of Amazon's program to fire their drivers through an automated email quite directly silences the voice of the drivers who are not given a chance to contest the decision.

2.4 CONCLUSION

The purpose of this chapter was to identify and analyze the ethical distance created by AI in decision-making and to contribute to the proposition of ethics of care as a form to counteract and mitigate the ethical implications of AI.

The discussion about ethics, distance, and technology is essential for AI ethics within business. Firms use algorithms without specific knowledge of their procedures, such as the COMPAS algorithm used in court sentencing to grant or not parole (Martin 2019b). There the "exclusivity of the individual is lost for the sake of technological palatability and optimization" (Nørskoy 2021). The distance and technologies de-skills firms' employees and make it easier to make decisions that could change a person's life, like decisions on an employee termination or acceptance to a university.

We examined how the abandonment of decision-making to AI has the ethical implication of moral distance. There we explained how moral distance arises from a proximity distance (of space, time, and culture) and from a bureaucratic distance (in the form of hierarchy, within complex processes, and principlism). We stated how these types of moral distance are presented in how AI works. Moreover, we argued that AI exacerbates the problem of moral distance. Finally, we have proposed the ethics of care as a moral framework to cover this issue, which implies understanding a moral responsibility to attend to interdependent relationships, vulnerabilities, context and circumstances, and to the consideration of what the other has to say.

Similar to all other ethical theories, the ethics of care can be considered unrealistic or not able to stop all moral problems or harms (Hamington 2019). We are arguing that the ethics of care is useful to analyze algorithmic decision-making given AI's negative impact on moral distancing. In this way, the ethics of care is useful given this particular context. In addition, the application of the ethics of care is to aid in the use of AI. The goal is to offer a mechanism to design and develop algorithmic decision-making tools that take into consideration the ethics of care.

REFERENCES

Allen C, Smit I, Wallach W (2005) Artificial morality: Top-down, bottom-up, and hybrid approaches. Ethics and Information Technology, 7(3):149-155

Bandura A (2002) Selective moral disengagement in the exercise of moral agency. Journal of Moral Education 31(2):101-119

Bauman Z (1989) Modernity and the Holocaust. Polity Press, Cambridge

Benkler Y (2019) Don't let industry write the rules for AI. Nature 569(7754):161-162.

Bia M, Kalika M (2007) Adopting an ICT code of conduct: An empirical study of organizational factors. Journal of Enterprise Information Management 20(4):432-446

Borgmann A (1987) Technology and the character of contemporary life: A philosophical inquiry. University of Chicago Press, Chicago

Brundage M, Avin S, Clark J, Toner H, Eckersley P, Garfinkel B et al (2018) The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. arXiv Preprint arXiv:1802.07228

Buolamwini J, Gebru T (2018) Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of the 1st Conference on Fairness, Accountability and Transparency PMLR 81:77-91

Burrell J (2016) How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society 3(1):1-12

Burton BK, Dunn CP (1996) Feminist ethics as moral grounding for stakeholder theory. Business Ethics Quarterly 6(2):133-147

Chatterjee DK (2003) Moral distance: introduction. The Monist 86(3):327-332

Coeckelbergh M (2013) Drones, information technology, and distance: mapping the moral epistemology of remote fighting. Ethics and Information Technology 15(2):87-98

Coeckelbergh M (2015) The tragedy of the master: automation, vulnerability, and distance. Ethics and Information Technology 17(3):219-229

Coeckelbergh M (2020) Artificial intelligence, responsibility attribution, and a relational justification of explainability. Science and Engineering Ethics 26(4):2051-2068

Cummings ML (2004) Creating moral buffers in weapon control interface design. IEEE Technology and Society Magazine 23(3):28-33

Dreyfus HL (2008) On the internet, 2nd edn. Routledge, London/New York

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

French W, Weis A (2000) An ethics of care or an ethics of justice. Journal of Business Ethics 27(1/2): 125-136

Gardiner SM (2003) The pure intergenerational problem. The Monist, 86(3):481-500

Gilbert B (2021) An Amazon driver said she nearly lost her house and had her car repossessed with her kids' Christmas presents inside after an algorithm suddenly fired her. Business Insider. https://www.businessinsider.com/amazon-driver-nearly-lost-house-when-an-algorithm-fired-her-2021-6. . Accessed 6 October 2021

Gilligan C (1982) In a different voice: Psychological theory and women's development. Harvard University Press, Cambridge, MA

Graham J, Meindl P, Beall E, Johnson KM, Zhang L (2016) Cultural differences in moral judgment and behavior, across and within societies. Current Opinion in Psychology 8:125-130

Grossman D (1995) On killing: The psychological cost of learning to kill in war and society. Little, Brown and Company, New York/Boston/London

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

Heidegger, M. 1977. The Question Concerning Technology and Other Essays . New York : Harper & Row.

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

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

Huber C., Munro I (2014) "Moral distance" in organizations: An inquiry into ethical violence in the works of Kafka. Journal of Business Ethics 124(2):259-269

Jonas H (1984) The imperative of responsibility: In search of an ethics for the technological age. University of Chicago press, Chicago

Jones C, Parker M, Ten Bos R (2005) For business ethics. Routledge, New York

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

Mac R (2021) Facebook apologizes after A.I. puts "primates" label on video of black men. The New York Times. https://www.nytimes.com/2021/09/03/technology/facebook-ai-race-primates.html. Accessed 20 September 2021

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

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

Mellema G (2003) Responsibility, taint, and ethical distance in business ethics. Journal of Business Ethics, 47(2):125-132

Mittelstadt B (2019) Principles alone cannot guarantee ethical AI. Nature Machine Intelligence 1:501-607

Nissenbaum H (1996) Accountability in a computerized society. Science and Engineering Ethics 2:25-42

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

Nørskoy M (2021) Robotification & ethical cleansing. AI & Society. https://link.springer.com/article/10.1007/s00146-021-01203-2. Accessed 10 September 2021

Prinsloo P (2020) Of 'black boxes' and algorithmic decision-making in (higher) education – A commentary. Big Data & Society January-June: 1-6

Prinsloo, P. and Slade, S. (2016). Student vulnerability, agency, and learning analytics: An exploration. Journal of Learning Analytics 3(1), 159-182.

Sander-Staudt M, Hamington M (2011) Introduction: Care ethics and business ethics. Applying Care Ethics to Business 34, VII

Segun ST (2021) Critically engaging the ethics of AI for a global audience. Ethics and Information Technology 23:99-105

Simola S (2010) Anti-corporate anger as a form of care-based moral agency. Journal of Business Ethics 94:255-269

Simola S (2015) Understanding moral courage through a feminist and developmental ethic of care. Journal of Business Ethics 130:29-44

Slade, S. and Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. American Behavioral Scientist 57(10), 1510-1529.

Reiter SA (1996) The kohlberg–gilligan controversy: Lessons for accounting ethics education. Critical Perspectives on Accounting 7:33-54

Rescher N (2003) By the standards of their day. The Monist 86(3):469-480

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

Thompson DF (1980) Moral responsibility of public officials - the problem of many hands. American Political Science Review 74(4):905-916

van de Poel I, Fahlquist JN, 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, Zwart SD (2015) Conclusions. From understanding to avoiding the problem of many hands. In: van de Poel I, Royakers L, Zwart SD (eds) Moral Responsibility and the Problem of Many Hands. Routledge, New York

Van den Bergh J, Deschoolmeester D (2010) Ethical decision making in ICT: discussing the impact of an ethical code of conduct. Communications of the IBIMA, Article ID 127497.

Weltzien H, Melé D (2009) Can an SME become a global corporate citizen? Evidence from a case study. Journal of Business Ethics 88:551-563

Wicks AC, Gilbert DR Jr, Freeman RE (1994) A feminist reinterpretation of the stakeholder concept. Business Ethics Quarterly 4(4):475-497

Zarsky T (2016) The trouble with algorithmic decisions: an analytic road map to examine efficiency and fairness in automated and opaque decision making. Science, Technology, & Human Values 41(1):118-132

Zyglidopoulos SC, Fleming PJ (2008) Ethical distance in corrupt firms: how do innocent bystanders become guilty perpetrators? Journal of Business Ethics 78:265-274

CHAPTER 3

WHAT THE ETHICS OF CARE IS NOT

> This conception of morality as concerned with the activity of care centers 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.9

The ethics of care appeared as a theory and term with Carol Gilligan in her pioneering study *In a different voice* in 198210. According to Gilligan, this new proposition regarding morality constructs moral decisions as something about "care and responsibility in relationships rather than as one of rights and rules … just as the conception of morality as justice ties development to the logic of equality and reciprocity."

Only four decades have passed, but a deep conversation has occurred around the ethics of care which has developed the notion of care within ethics into a broader approach in its designation and application. As it happens with other disciplines, scholars use a variety of terms in referring to this moral theory, *care*, *caring*, the *ethic*(*s*) *of care*, or *care ethics*. Also, contemporary perspectives regarding this theory emerged from various disciplines, including but not limited to psychology, moral philosophy, bioethics, political theory, education, or business ethics.11 Ethicists delineate the notion of care in

9 C. Gilligan, *In a different voice* (Harvard University Press, Cambridge, MA., 1982).
10 C. Gilligan, *In a different voice* (Harvard University Press, Cambridge, MA., 1982). Some scholars argue that the formulation of ethics of care is before Gilligan. However, she was the first to propose the term and the one that started the conversation about it.
11 This list also includes some of the most relevant works in the field. From psychology: C. Gilligan, *In a different voice* (Harvard University Press, Cambridge, MA., 1982). From moral philosophy: A. Baier, *What do women want in a moral theory?*, "Nous" 19 (1985) 53-63.; A. Baier, *Moral prejudices:essays on ethics* (Harvard University Press, Cambridge, MA, 1994). From bioethics: J. Harbinson, *Gilligan: a voice for nursing?* "Journal of medical ethics" 18 (1992) 202-205.; R. Gillon, *Caring, men and women, nurses and doctors, and health care ethics*, "Journal of medical ethics" 18/4 (1992) 171-172.; For political theory: J.

numerous ways: as a set of dispositional attitudes, as a practice or something that must be done on a direct level (a face-to-face interaction), or as labor that is inherently relational.12

Scholars in the field have developed a broader understanding of what an ethic of care is. Nevertheless, some concepts cannot be reduced to definitions,13 and the ethics of care appears to be one of them. The first generation of scholars in the field, including Gilligan and Noddings, did not offer a definition of the ethics of care. Even prominent care ethicists, such as Virginia Held, refused openly to offer a definition, and she argued: "What *is* care? What do we mean by the term 'care? Can we define it in anything like a precise way? There is not yet anything close to agreement among those writing on care on what exactly we should take the meaning of this term to be."14 Hence, some theorists argue that is better to leave the definition of the ethics of care ambiguous, "care ethicists seem to suspect something important would be lost in the assertion of a slogan, so they do not attempt to provide a clear statement of *the* normative core of the theory"15 But some others, like Stephanie Collins, find the ambiguity problematic16 and the cause of misunderstandings.

Recent academics have sharpened their definitions. Daniel Engster, one of the prominent scholars in the field, addressed the issue of how the ethics of care is usually defined in ways that give rise to ambiguous interpretations of the theory. In that context,

---

H. Tronto, *Moral Boundaries: a political argument for an ethic of care* (Routledfe, London, UK).; D. Engster, *The Heart of justice. Care ethics and political theory* (Oxford University Press, New York, 2007). From education: N. Noddings, *Caring: A relational approach to ethics and moral education* (University of California Press, Berkeley, CA, 2013).; N. Noddings, *The Challenge to care in schools: an alternative approach to education* (University of California Press, Berkeley, CA). From business ethics: M. Hamington and M. Sander-Staudt (Eds.), *Applying care ethics to business* (Springer, Oxford, 2011).

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

13 Notes from the class "Forms of representation of power" of professor Jaume Aurell in the Master's degree in Organizational Governance and Culture, University of Navarra, 2016)

14 V. Held, *The ethics of care: Personal, political, and global* (Oxford University Press, New York, 2006).

15 S. Collings, *The Core of Care Ethics* (Palgrave Macmillan, UK, 2015).

16 S. Collings, *The Core of Care Ethics* (Palgrave Macmillan, UK, 2015).

Engster focused on the core aims of consensus of previous literature and offered a definition of what *care* is within the ethics of care:

> Everything we do directly to help individuals to meet their vital biological needs, 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.17

Some years after his delineation of what care is, 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.18

Still, the notion of care in ethics is commonly misunderstood,19 and this may come from the contextualized moral point of view that it proposes. The ethics of care bolsters in the understanding that responsibility only makes sense in a web of relationships, and in there, moral discourse starts to be about interdependence, vulnerability, rooted in a specific circumstance where the generalized other is an individual with a history, and with a story to tell, with their own voice. This is what the ethics of care is about. However, some misinterpretations of its propositions may lead to the rejection of the theory from scholars in the field of ethics. Perhaps, more than a rejection, is the typical reaction that occurs in the face of breakthrough, in the face of

17 D. Engster, *The Heart of justice. Care ethics and political theory* (Oxford University Press, New York, 2007).
18 D. Engster, *Care ethics and stakeholder theory*, in M. Hamington and M. Sander-Staudt (Eds.), *Applying care ethics to business* (Springer, Oxford, 2011) 93-110.
19 P. Allmark, *Can there be an ethics of care?*, "Journal of medical ethics" 21 (1995) 19-24.; J. Paley, *Caring as a slave morality: Nietzschean themes in nursing ethics*, "Journal of Advanced Nursing" 40/1 (2002) 25-35.; S. D. Edwards, *Three versions of an ethics of care*, "Nursing Philosophy" 10 (2009) 231-340.

what comes to challenge our common and traditional understanding of how things are and work.

Regardless of its intention, these misunderstandings lead skeptics of the ethics of care to allege, among other things, that the theory leads to ambiguity,20 which can fall into partiality, excessive sentimentalism, or feelings of pity. Moreover, some of the common misinterpretations allege that the theory completely rejects a vision close to justice or even that the theory does not say anything new and therefore it would make no sense to recognize it as such. In this sense, Margaret Olivia Little argued that the ethics of care is not a theory but a moral *orientation* and that it's "contribution to moral theory is best seen as stances from which to do theory, rather than as constituting ready-made" theory itself.21

In order to help to avoid some of the most common misunderstandings of the ethics of care, the purpose of this paper is to identify what the ethics of care is not. This proposition is an attempt to clarify the implications of the moral theory without narrowing the scope of it, but with the identification of its boundaries.

For this purpose, I will analyze four main (incorrect) claims that the ethics of care is not. First, the ethics of care is not altruism. Its call to help to meet the needs of those who cannot meet them by themselves does not entail a self-sacrifice where the one who cares is affected, but it also asks for self-care. Second, the ethics of care is not about partiality, is not about feeling sorry or doing someone a favor but a demand for a holistic understanding of responsibility. Third, the ethics of care is not only about women (even if it has its roots in a feminist conversation). And finally, the ethics of care is not a part

20 P. Allmark, *Can there be an ethics of care?*, "Journal of medical ethics" 21 (1995) 19-24.; J. Paley, *Caring as a slave morality: Nietzschean themes in nursing ethics*, "Journal of Advanced Nursing" 40/1 (2002) 25-35.; J. Paley, *commentary: Care tactics – arguments, absences and assumptions in relational ethics*, "Nursing Ethics" 18/2 (2011) 243-254.
21 M. O. Little, *Care: From Theory to Orientation and Back*, "Journal of Medicine and Philosophy" 23/2 (1998) 190-209.

of virtue ethics, rather it starts a new dialogue, has new propositions and an establish contribution to the ethics field. In each part, I will refer to what the ethics of care *is* and what this view of morality proposes, but with a focus and emphasis on what is *not*. This approach appears to be the way to avoid some of the most common misinterpretations when talking about the ethics of care. I developed this contribution based on the point of consensus among scholars in the field. Hence, the literature on the ethics of care has already addressed these issues separately. Either directly or indirectly, academics in this branch have defended and clarified that these are misinterpretations of the proposals.

3.1. THE ETHICS OF CARE IS NOT ALTRUISM

> The ideal of care is thus an activity of relationship, of seeing and responding to need, taking care of the world by sustaining the web of connection so that no one is left alone.22

One of the core aims of the ethics of care is the reference to those who have no voice, to the marginalized or vulnerable. Ethicists in this field constantly refer to the responsibility of the individual to help others to meet their needs. However, *care* within ethics should be distinguished from *personal service*, "the former involves meeting the needs of those who are unable to meet such needs themselves, the latter involves meeting the needs for others who could meet such needs themselves."23

An important critique of the ethics of care arises from the fact that one cannot practically care for everyone's needs. This idea addresses the factual impossibility of individuals to care for all the problems, injustices, or atrocities that can see or recognize all over the world. That criticism comes from the misunderstanding of care as altruism.

22 C. Gilligan, *In a different voice* (Harvard University Press, Cambridge, MA., 1982).
23 M. Hamington and M. Sander-Staudt (Eds.), *Applying care ethics to business* (Springer, Oxford, 2011). See also Bubeck, D. *Care, gender and justice* (Clarendon Press, Oxford, 1985).

Nel Noddings addressed this misinterpretation with her distinction between *caring-for* and *caring-about*:

> Caring-for describes an encounter or set of encounters characterized by direct attention and response. It requires the establishment of a caring relation, person-to-person contact of some sort. Caring-about expresses some concern but does not guarantee a response to one who needs care. We are all familiar with an array of scenarios that might be called "caring-about." I might, for example, care about civilians living in fear during civil strife in, say, Syria, but I may not follow up on my expressed concern. Or I may follow up with a small gift to a charitable organization. Edging closer to caring-for, I may check on the credentials of the organization to find out how my contribution is spent. The point is that we cannot care-for everyone; we are limited by time, resources, and space. My comment that we have no obligation to care-for the starving children of Africa generated outrage in many readers. But we cannot be obligated to do the impossible, and it is clearly impossible to establish a caring relation with everyone in the world.24

This distinction of the philosopher Nel Noddings appears in the preface to the 2013 edition of her book *Caring,* originally published in 1984 and which is usually presented as one of the two seminal works in the ethics of care (along with Gilligan's 1982 book). The issue of equating the ethics of care with altruism is that the theory then loses its effectiveness as an ethical system in most contexts. Then it's not possible to reference the ethics of care as a normative theory of what people ought to do since we do not have a normative grounding of altruism: Altruism is an option but not a positive

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

obligation to others. But we can certainly refer to the propositions of care as something that we are responsible for, as the case of the relation employer-employee, or enterprise-community, or the so long referred relation parent-child. Here is essential to point out the idea of Gilligan when talking about relationships: "the logic underlying an ethic of care is a psychological logic of relationships, which contrast with the formal logic of fairness that informs the justice approach."25 Hence, one relationship to the other is in part what determines the outcome and not a well-intentioned will to do good.

The ethics of care is not altruism also because the theory does not imply or argue for self-sacrifice or the idea to help the other even at the cost of limiting or removing one's own. Victoria Davion argued that Noddings proposed a one-caring relationship model which appeared inappropriate for mature relationships that should be based on reciprocity and not one-way care.26 Other critics suggested that the ethics of care undermines women's autonomy while reinforcing traditional roles where women stayed at home and are the ones in charge of caring activities.27 This critic comes again with the idea of one-way care and the ethics of care as altruism.

Conversely, ethicists in the field often refer to the need to care first for oneself to be able to help the other, just as Virginia Held best summarizes in this statement:

> Care is not a kind of charity or benevolence, it is not based primarily on altruism. It rejects the reduction of our choices to egoism versus altruism. It opposes the misconstrual of innumerable situations as being such that promoting the interests of the self or the interests of the other are the only alternatives. Care aims at the well-being of both providers and recipients of care … But it would not see the moral requirements as calling primarily

25 C. Gilligan, *In a different voice* (Harvard University Press, Cambridge, MA., 1982).
26 V. Davion, *Autonomy, Integrity, and Care*, "Social Theory and Practice" 19/2 (1993) 161-182.
27 J. Keller, *Autonomy, Relationality, and Feminist Ethics*, "Hypatia" 12/2 (1995) 128-133.

for charity or self-sacrifice. It would understand them as part of a mutual commitment to mutual well-being.28

There Held addressed the common criticism that the ethics of care could imply leaving behind the urge to care for themselves for those who engage in relationships of care. Also, Maurice Hamington and Daniel Engster definitions of care addressed the moral theory as considering the relevance of self-care.29 Moreover, Gilligan has come to the attention of this misunderstanding, and she has stated that "without a voice, there is no relationship, only the chimera of relationship," and that "to have a voice means to be present, not absent with oneself and with others. The sacrifice of voice for the sake of relationships [is] psychologically incoherent."30 Within this line, Gilligan questioned the idea that if is good to be responsive to the needs of others, why is it then selfish to respond to the needs of oneself? Hence, the ethics of care argues for a balance between care for other and for self, and also an equilibrium between care for distant and close individuals.31 In the same conversation, other scholars have defended that even that care is usually related to kindness and gentleness, the act of care may entail in many situations a need for anger and to defend what should be done in a challenging and controversial way.32 To help in this conversation, Tove Pettersen proposed the idea of *mature care* which conceptualizes the notion of care as relational and not as a mono-directional activity.33

28 V. Held, *Taking responsibility for global poverty*, "Journal of Social Philosophy" 49/3 (Fall 2018) 393–414.
29 D. Engster, *The Heart of justice. Care ethics and political theory* (Oxford University Press, New York, 2007); M. Hamington, *Embodied care: Jane Addams, Maurice Merleau-Ponty and Feminist Ethics* (Rowman & Littlefield, New York, 2004).
30 C. Gilligan, *Revisiting "in a different voice,"* "The Harbinger" 39/1 (2015) 19-28.
31 V. Held, *Taking responsibility for global poverty*, "Journal of Social Philosophy" 49/3 (Fall 2018) 393–414.
32 S. Simola, *Anti-corporate anger as a form of care-based moral agency*, "Journal of Business Ethics" 94 (2010) 255-269.
33 T. Pettersen, *Conceptions of care: altruism, feminism, and mature care*, "Hypatia" 27/2 (2012) 266-389.

3.2. THE ETHICS OF CARE IS NOT ABOUT PARTIALITY

Another misunderstanding regarding the ethics of care is to frame it as a moral theory that promotes partiality or the call to do someone a favor. The kind of partiality that I refer to here is the one guided by feelings of pity (sometimes at the expense of others) . Instead, the propositions of the ethics of care are not about breaking principles – doing someone a favor out of pity, because you feel sorry for them – it questions the principles we develop. When talking about the ethics of care and the problem of partiality, Patrick Boleyn-Fitzgerald explain the difference between compassion and pity and propose to understand the moral theory in line with compassion rather than pity.

> Compassion is a combination of empathy, benevolence, and equanimity. In other words, compassion is an empathetic understanding of another's suffering, an equanimous holding of any suffering or risk of suffering that cannot be relieved, and a determination to relieve any suffering that can be relieved. Pity is also an emotional and benevolent response to suffering, but it does not involve equanimity.34

Given the continuous reference of scholars in the ethics of care to context and circumstances in ethical decision making, John Paley complained that "it is not at all clear how care ethics can favor equal concern for all those affected, and at the same time give precedence to the most proximate," there they explained that ethicists in the field does not appear to be "entirely consistent about the nature of partiality."35 Also, Peter Allmark in his article *Can there be an ethics of care?* Equate the ethics of care to partiality, an

34 P. Boleyn-Fitzgerald, *Care and the problem of pity*, "Bioethics" 17/1 (2003) 1-20
35 J. Paley, *commentary: Care tactics – arguments, absences and assumptions in relational ethics*, "Nursing Ethics" 18/2 (2011) 243-254.; See also, S. D. Edwards, *Three versions of an ethics of care*, "Nursing Philosophy" 10 (2009) 231-340.

even uses bot terms as synonyms.36 In this line, Steven D. Edwards developed his critique:

> Although, as proponents of an ethic of care point out, our emotional response would lead us to give priority to the interests of our loved ones, and ourselves, this kind of 'partialism' looks problematic as an approach to ethics. This is because partialism, seemingly arbitrarily, attaches greater weight to the protection of one's own interests above protection of the interests of others – especially those who are moral strangers. Critics complain that no such partialist approach to ethics can be plausible.37

However, the quid here is that equating the propositions of the ethics of care to a matter of partiality may entail doing someone a favor out of feelings of pity. Indeed, the ethics of care argue for a contextualized point of view in ethical decision-making. This focus entails the constant reference to context and circumstances, the current relationships of interdependence and vulnerabilities (note that in many cases the dependence or vulnerability could disappear as in the case of someone with an illness who latter recovers). These emphasis in the ethics of care further develop into a position where principles (or principle base ethics) are not sufficient, because they are developed for the generalized other rather than for a concrete individual. Also, principles or ethical guidelines may be necessary but not enough for certain circumstances. Hence, the idea of insufficiency does not entail the elimination of principles or ethical guidelines for scholars in the ethics of care, the next statement of Nel Noddings is an example of the said:

36 P. Allmark, *Can there be an ethics of care?*, "Journal of medical ethics" 21 (1995) 19-24.
37 S. D. Edwards, *Three versions of an ethics of care*, "Nursing Philosophy" 10 (2009) 231-340.

> Care theorists, like virtue ethicists, put limited faith in principles. We not disdain principles. We recognize that principles—for example, an injunction against lying—help to keep daily life running smoothly. We learn the rule and commit ourselves to it, and for the most part we would not consider breaking it. However, when a real conflict arises, the principle is of little help. We have to dig behind the principle to see what deeper value has engendered it. Most of us would not consider adopting the rule as an absolute; we would not join Kant in refusing to lie to a would-be murderer even to save the life of the victim. Instead, we ask who might be hurt, who might be helped by our lie.38

Hence, following the propositions of the ethics of care, one can say that traditional moral theories, who narrow their understanding of morality to principles, limit responsibility to the justification of actions by appealing to previously stablished codes of conduct.39 What the ethics of care proposes is the opposite to this view, while bringing to the conversation the idea that an ideal individual in an ideal circumstance does not exist, but every case is different since every human being is unique.

The call to leave behind principles when they are not sufficient, or in cases when a blind attachment to principles may lead to harm, should not be understood as doing someone a favor, this idea fundamentally challenges policies and rules but does not bend for certain people. Hence, "it is not, as Shakespeare's Portia demanded, mercy that is to season justice, but a less authoritarian humanitarian supplement, a felt concern for the

38 N. Noddings, *Caring: A relational approach to ethics and moral education* (University of California Press, Berkeley, CA, 2013)
39 T. Petterson, *Comprehending Care: Problems and Possibilities in the Ethics of Care* (Rowman & Littlefield, Lanham, MD, 2008).

good of other and for community with them."40 Here it may serve to shed light the difference between empathy and sympathy clarified by schooler Daryl Kohen:

> While sympathy may sometimes be an appropriate response, it is also problematic because it often means that the agent is feeling sorry for someone else. Insofar as I operate from a position of superiority and pity my friend, I am not entering into her feelings and cannot be said to be thinking empathically. In other cases, an agent may experience a feeling of sympathy quite independently of what others are feeling. The death of a great artist may sadden me, irrespective of what the artist's relatives feel. Since sympathy often operates independently of others' concerns, it cannot be relied upon to provide any real insight into their life experiences."41

Let's illustrate all the said in an example. Suppose that within the COVID-19 pandemic a company decides to fire several employees to save money. The firm decide to fire those (let's say) ten with the lower performance according to the economic benefit they have meant for the company. A misunderstanding of the ethics of care, one where paritality is assumed, would suggest that the firm should make exceptions for those whose circumstances are difficult. Suppose that one of the employees with lower rates is a parent of two little kids, and who has been in charge of the care of the girls during homebound, having no other option than focusing on the kids' needs. A misunderstanding of the ethics of care as promoting partiality would suggest making an exception for that worker. But the ethics of care is not based on favoring someone over others. Instead, the ethics of care would question the policies or rules as not being developed with a contextualized understanding of the voices and vulnerabilities of the workers. Rather than making

40 A. Baier, *Moral prejudices:essays on ethics* (Harvard University Press, Cambridge, MA, 1994).
41 D. Kohen, *Rethinking Feminist Ethics: Care, Trust, and Empathy* (Routledge, NYC, 1998).

exceptions, the ethics of care would suggest creating policies in time of the pandemic to accommodate the lived experiences of their workers. This may require saving money by not paying dividends to shareholders or saving money by asking employees to voluntarily work less if they can afford it.

Hence, we are not talking about feeling sorry for any worker or doing someone a favor but comprehending the situation holistically. The misunderstanding of the propositions of the ethics of care as to bend the rules and principles for certain people based on a partiality perspective is a common complaint to contextualized theories. Since there isn't only one answer, people error to think that the proposition is that *anything goes*, but is questioning the principles we develop.

3.3. THE ETHICS OF CARE IS NOT ONLY ABOUT WOMEN

Gilligan presented an ethic of care as an alternative to an ethic of justice, but *opposed* as different, not against, not instead of. Her work was a response to the studies and propositions of her advisor Lawrence Kohlberg42. Kohlberg proposed three levels of moral development and six stages (two stages within each level)43. According to Gilligan, the main problem of these six stages is that they are empirically based on a study of eighty-four boys that Kohlberg studied during a period of over twenty years. Gilligan defended that Kohlberg's claims of universality for his theory of moral development are not real since those groups excluded in his original sample hardly reach the higher stages of his sequence. What Gilligan noticed was that judgments of woman appeared to

42 L. Kohlberg, *The development of modes of thinking and choices in year 10 to* 16 (Ph.D. Diss., University of Chicago, 1958). L. Kohlberg, *Essays on Moral Development* (Harper & Row, New York, 1981).
43 The first was the *Preconventional level*, with stage 1. The punishment-and-obedience orientation, and stage 2. The instrumental-relativist orientation. The second one was the *Conventional level*, with stage 3. The interpersonal concordance "good boy – nice girl", and stage 4. The "law and order" orientation. Finally, the third level was the *Postconventional, autonomous, or principled level*. In this last level was the stage 5. The social-contract, legalistic orientation, generally with utilitarian overtones, and stage 6. The universal-ethical-principle orientation. See L. Kohlber and R. H. Hersh, *Moral Development: A review of the theory*, "Theory Into Practice" 16/2 (1977) 53-59.

exemplify the third stage of the sequence, when morality is conceived in interpersonal terms and to be good is equated with helping and pleasing the other.

> This conception of goodness is considered by Kohlberg and Kramer (1969) to be functional in the lives or mature women insofar as their lives take place in the home. Kohlberg and Kramer imply that only if women enter the traditional arena of male activity will they recognize the inadequacy of this moral perspective and progress like men toward higher stages where relationships are subordinated to rules (stage four) and rules to universal principles of justice (stages five and six).44

Kohlberg was a proponent of justice approaches, and his theory is one of the prominent theories of moral reasoning. However, for Gilligan, his articulation of a six-stage sequence fails to listen to diverse approaches and ends up rating as deficient to those who do not fit the pattern of his sample. There, Gilligan proposed an *ethic of care* as an alternative to an *ethic of justice*, the latter in reference to Kohlberg's proposition and the sixth stage of the universal ethical-principle orientation.

Hence, the ethics of care has its roots in a feminist conversation since Gilligan was defending to hear the positions of women. Also, his book *In a different voice* has as subtitle *Psychological theory and women's development*. However, Gilligan has long argued that she was referring to a *different* voice in its broader sense, not only referring to women. Also, Nel Noddings changed the subtitle of her book *Caring* in the edition of 2013 from a *feminine* approach to *A relational approach to ethics & moral education*. This is how she explained the need to change it:

> Hardly anyone has reacted positively to the word feminine here. In using it, I wanted to acknowledge the roots of caring in women's experience …

44 C. Gilligan, *In a different voice* (Harvard University Press, Cambridge, MA., 1982

> I think critics are right, however, to point out that the connotations of "feminine" are off-putting and do not capture what I intended to convey. Relational is a better word. Virtually all care theorists make the relation more fundamental than the individual.

Here I stated not *only* related to women because one cannot deny that the conversation about an ethic of care has its roots in a feminist context, and that the pioneering ethicist within the ethics of care continually followed the conversation from an overtly feminist point of view. Moreover, critics stated that the proposition of pioneering scholars within the ethics of care "seems to eschew commitment to a formal principle of justice,"45 referring to Gilligan, and that the "version of feminist care ethics professes to be incompatible with Kant's ethics,"46 referring to Noddings.

But nowadays, the theory has developed in many different aspects into a broader understanding of the ethics of care as a comprehensive approach (complementing within other theories) and not only related to women. In fact, scholars can refer to this moral theory without reference or direct implication to women. Daniel Engster argued that objections of care ethicists to justice frameworks are weak and proposed to reframe the ethics of care in terms of a theory of justice, "in order to make it more accessible to readers outside the field and more applicable to practical politics," and he defended that with that frame he removes the theory "from the feminist context in which it has developed."47The philosopher Annette Baier explains this approach in this statement of her book *Moral Prejudices* in 1994:

45 S. D. Edwards, *Three versions of an ethics of care*, "Nursing Philosophy" 10 (2009) 231-340.
46 V. S. Wike, *Defending Kant Against Noddings' Care Ethics Critique*, Kant Studies Online Ltd. (2011) 1-26. Last time retrieved December 2021, https://kantstudiesonline.net/uploads/files/WikeVictoria00611.pdf
47 D. Engster, *The Heart of justice. Care ethics and political theory* (Oxford University Press, New York, 2007).

> It is clear, I think, that the best moral theory has to be a cooperative product of woman and men, has to harmonize justice and care. The morality it theorizes about is after all for all persons, for men and for woman, and will need their combined insights.48

However, scholars from other disciplines continue to frame the theory as only related to women. The problem of this misinterpretation of the theory is that it narrows the scope of the same and make it impossible to reach the core point of this moral theory that entails not only women but every affected part of the equation, especially those who given its circumstances cannot defend their own needs. "This enlarged conceptual framework provides a new way of listening to differences not only between but also within the thinking of women and men."49 Understanding the ethics of care as a theory only related to women, one can ignore the fact that "it is concerned with how, in general, we should meet and treat one another—with how to establish, maintain, and enhance caring relations"50 regardless the feminist debate.

48 A. Baier, *Moral prejudices:essays on ethics* (Harvard University Press, Cambridge, MA, 1994).
49 C. Gilligan, *Mapping the moral domain: new images of self in relationship*, "CrossCurrents" 39/1 (1989) 50-63.
50 N. Noddings, *Caring: A relational approach to ethics and moral education* (University of California Press, Berkeley, CA, 2013).

3.4. THE ETHICS OF CARE IS NOT A PART OF VIRTUE ETHICS

Ethicists have broadly debated about whether the ethics of care should be understood as a part of virtue ethics or as an independent framework. Although this moral framework is related to other theories and may be especially compatible with the propositions of virtue ethicist, scholars that work in the field of ethics of care openly differentiate themselves from the work within virtue ethics.

> "The ethics of care is a distinct moral theory or approach to moral theorizing, not concern that can be added on to or included within other more established approaches, such as those of Kantian moral theory, utilitarianism, or virtue ethics. The latter is more controversial claim, since there are similarities between the ethics of care and virtue ethics. But in its focus on relationships rather than on the dispositions of individuals, the ethics of care is, I argue, distinct."51

Since its inception, the ethics of care has opened a different focus to comprehend morality. The idea of Held argues that ethicists in the field should be reluctant to equate the propositions of the ethics of care as a matter of virtue, "because this runs the risk of losing sight of it as work."52

Daniel Engster differentiated care as a virtue, and as practice, the first one focuses on the "inner traits, dispositions, and motivations of the caring person," But when care is defined as a practice, the focus is on external actions and consequences53. The first one

51 V. Held, *The ethics of care: Personal, political, and global* (Oxford University Press, New York, 2006).
52 V. Held, *The ethics of care: Personal, political, and global* (Oxford University Press, New York, 2006).
53 D. Engster, *The Heart of justice. Care ethics and political theory* (Oxford University Press, New York, 2007).

is clearly related to the virtue ethics approach, which implies that both theories are not completely opposed. However, the two have nevertheless different emphases.

Within the ethics of care, one asks specific and different questions. Virtue ethicists tried to answer to how humans can live well in society and flourish? Or which are the dispositions and character that a good person should develop. On the other hand, with the focus of an ethic of care in any situation, one should ask which voices are being silenced? Which interdependent relationships should be attended? Which needs are being oppressed or ignored? Also, does the decision that is being made disregard implications of context and circumstances? Does it neglect vulnerability or emotions? Besides the theory, like many others, bolsters in the broad spectrum of morality, the ethics of care entails a new focus and emphasis, a new conversation in it.

3.5. CONCLUSION

The purpose of this paper was to identify and analyze what the ethics of care is not, as an attempt to clarify the implications of the moral theory without narrowing the scope of the notion but with the clarification of its boundaries to clear up misconceptions about this moral theory. To conquer this purpose, I defend first that the ethics of care is not altruism, since it does not propose one-way care, but appeals to the relevance of self-care. Hence the ethics of care is a relational theory and non a mono-directional proposal. Second, the ethics of care is not about partiality, nor about feeling sorry within ethical decision-making, but a contextualized theory that states that there is not only one answer (not either that anything goes) but a challenge to the principles we develop to give response to each circumstance. Third, the ethics of care is not only about women, is a comprehensive theory, in line with other moral theories, that makes a universal proposal,

not centered on the conversation of women. Finally, the ethics of care is not virtue ethics, nor a part of it, since it starts a new conversation, asks new questions, and proposes new concerns within the field of ethics.

REFERENCES

A. Baier, *What do women want in a moral theory?*, "Nous" 19 (1985) 53-63.D. Bubeck, *Care, gender and justice* (Clarendon Press, Oxford, 1985).

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

C. Gilligan, *Mapping the moral domain: new images of self in relationship*, "CrossCurrents" 39/1 (1989) 50-63.

C. Gilligan, *Revisiting "in a different voice,"* "The Harbinger" 39/1 (2015) 19-28.

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

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

J. Harbinson, *Gilligan: a voice for nursing?* "Journal of medical ethics" 18 (1992) 202-205.

J. H. Tronto, *Moral Boundaries: a political argument for an ethic of care* (Routledfe, London, UK).

L. Kohlberg, *The development of modes of thinking and choices in year 10 to* 16 (Ph.D. Diss., University of Chicago, 1958).

L. Kohlberg, *Essays on Moral Development* (Harper & Row, New York, 1981).

L. Kohlberg and R. H. Hersh, *Moral Development: A review of the theory*, "Theory Into Practice" 16/2 (1977) 53-59.

M. Hamington, *Embodied care: Jane Addams, Maurice Merleau-Ponty and Feminist Ethics* (Rowman & Littlefield, New York, 2004).

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

M. O. Little, *Care: From Theory to Orientation and Back*, "Journal of Medicine and Philosophy" 23/2 (1998) 190-209.

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

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

N. Noddings, *The Challenge to care in schools: an alternative approach to education* (University of California Press, Berkeley, CA).

P. Allmark, *Can there be an ethics of care?*, "Journal of medical ethics" 21 (1995) 19-24.

P. Boleyn-Fitzgerald, *Care and the problem of pity*, "Bioethics" 17/1 (2003) 1-20

J. Keller, *Autonomy, Relationality, and Feminist Ethics*, "Hypatia" 12/2 (1995) 128-133.

J. Paley, *commentary: Care tactics – arguments, absences and assumptions in relational ethics*, "Nursing Ethics" 18/2 (2011) 243-254.

R. Gillon, *Caring, men and women, nurses and doctors, and health care ethics*, "Journal of medical ethics" 18/4 (1992) 171-172.

S. Collings, *The Core of Care Ethics* (Palgrave Macmillan, UK, 2015).

S. Simola, *Anti-corporate anger as a form of care-based moral agency*, "Journal of Business Ethics" 94 (2010) 255-269.

T. Monchiski, *Education in hope: critical pedagogies and the ethic of care (Peter Lang, New York, 2010).*

T. Petterson, *Comprehending Care: Problems and Possibilities in the Ethics of Care* (Rowman & Littlefield, Lanham, MD, 2008).

T. Pettersen, *Conceptions of care: altruism, feminism, and mature care*, "Hypatia", 27/2 (2012) 266-389.

V. Davion, *Autonomy, Integrity, and Care*, "Social Theory and Practice" 19/2 (1993) 161-182.

V. Held, *Taking responsibility for global poverty*, "Journal of Social Philosophy" 49/3 (Fall 2018) 393–414.

V. Held, *The ethics of care: Personal, political, and global* (Oxford University Press, New York, 2006).

V. S. Wike, *Defending Kant Against Noddings' Care Ethics Critique*, Kant Studies Online Ltd. (2011) 1-26. Last time retrieved December 2021, https://kantstudiesonline.net/uploads/files/WikeVictoria00611.pd

CHAPTER 4

THE ETHICS OF CARE IN THE ERA OF ARTIFICIAL INTELLIGENCE

In Philosophy and Business Ethics. *Organizations, CSR, and Moral Practice*.

G. Faldetta, Mollona, E., Pellegrini, M. M. (Eds.) *Philosophy for Business Ethics*. Palgrave Macmillan. ISBN-13: 978-3030971052

Carolina Villegas Galaviz

José Luis Fernández Fernández

ABSTRACT

The arrival of AI in management decision-making promises possibilities for many enhancements in efficiency and speed. However, the promises imply significant challenges, like ethical gaps to address. Hence, in this article, we propose ethics of care as moral grounding for the AI era in management. The article's structure is as follows: first, we present the context of the arrival of AI and its ethical implications. Second, we present the ethics of care, its main premises, and its application in business ethics and stakeholder theory. Finally, we approach ethical problems in management decision-making from a care ethics perspective, and then we propose a principle for companies. Finally, we offer some conclusions.

Keywords: Business ethics, AI Ethics, Stakeholders Theory, Ethics of Care

4.1. INTRODUCTION

Since the pre-industrial society, technology has changed the way that humans work (Liker et al. 1999; Aronowitz and Difazio 1996). For decades, some technologies' arrival has changed the nature of business (Hill and Rothaermel 2003): the steam engine, electricity, and ICT are some examples. We are now at the beginning of the Fourth Industrial Revolution (Schwab 2016; see also Ghobakhloo 2020). Nowadays, technology companies mark the rhythm of business and technology innovations as Artificial Intelligence (AI), Big Data, and Business Analytics mark these enterprises' rhythm (Wiener et al. 2020). In this scenario, the accelerated pace of science and technology leaves many ethical gaps to address (Jonas 1984; 1985): automation and unemployment (Dodel and Mesch 2020; Kim and Scheller-Wolf 2019; Wright and Schultz 2018), AI and decision-making (Cervantes et al. 2016; Robbins and Wallance 2007), and the like. There are many studies on business ethics and many journals dedicated to the compilation of studies on this subject. Moreover, the implications of technology in business ethics have been studied from different points of view (Buchholz and Rosenthal 2002; Davies 2002; Peace et al. 2002; Yuthas and Dillard 1999), and different philosophers have studied the philosophical side of technology and its implications in society (Heidegger, 1977; Habermas 1968), and in work (Marx 1932; Bell 1973).

However, fewer studies have covered the morality and values of technology development in business and the moral consequences of introducing innovations such as Business Analytics and AI can have in society. In this line, one of the main problems of ethics is the accelerated way technology advances, as change is an essential part of technology (Jonas 1984; 1985). Hence, an ethic for the Fourth Industrial Revolution must continuously adapt to change, technology development, and social evolution. The latter

means adapting to changing values (Van de Poel 2018). Since society changes, the relevance of moral values can change (Van den Hoven et al. 2015). With the application of new technology, new values and moral problems arise, and with them appears the need for an adapted moral code and conception of moral responsibility. In this scenario, the aim of this chapter is to propose the theory of ethics of care to contribute to the well-design, development, and deployment of AI for its use in management and business. This aim will be materialized in the proposition of a care-based principle for AI management decision-making process while considering all stakeholders' needs.

The chapter's structure is as follows: first, we present the context of the arrival of AI to firms and its ethical implications. Secondly, we present the ethics of care, its main premises, and its application in business ethics and stakeholder theory. Finally, we approach ethical problems in management decision-making from a care ethics perspective, and then we propose a principle for firms. We ended up with some conclusions and future research.

4.2. CONTEXT

In the White Paper *On Artificial Intelligence - A European approach to excellence and trust*, recently published in February 2020, the European Commission defines AI as a collection of technologies that combines data, algorithms, and computing power. In this research, we will focus on the ethical implication of AI in its relationship with companies, leaving aside the moral aspects of responsible innovation of AI as a separate discipline. In its informative document, the European institution presents AI applications as high-risk tools, although it is specified that they can bring significant benefits to nations. We investigated both sides of AI, on the one side, the risks that it can imply, and, on the other side, we centered our efforts to make a contribution in which the ethics of care can help

to ameliorate the *dark* side of AI or the fact that AI is usually associated with lack of privacy, problems with algorithms bias (i.e., socioeconomic inequality, racism), and the like.

Despite being a recent topic, several high-quality journals have published articles that address specific issues about AI ethics in business organizations, i. e. the Journal of Business Ethics (February 2022) and the California Management Review (August 2019) edited two journal special issues. Within this scheme appeared the so-called *machine ethics* (Anderson and Anderson, 2011), which study the moral issues that arise with the implementation of AI technologies, such as decision-making by autonomous systems (Awad et al. 2019; Shank et al. 2019). Some scholars in this branch of ethics presuppose that autonomous systems work with algorithms loaded with a neutral moral intent, but others defend the opposite (see Martin 2019, where the author present algorithms as value laden). However, for the machine ethics field, the objective is to find a consensus to create a moral guide for machines (Anderson et al. 2019).

In the search for moral principles that guide machine ethics, several moral guides on AI have been proposed (Anderson et al. 2019; Awad et al. 2019; Cervantes et al. 2016; Floridi 2018; Floridi and Tadeo 2016). Most of the models allude to what they identify as universal principles of all ethical agents, such as not killing, not lying, or not stealing (Cervantes et al. 2019). That is, the respect of agents other than the self. Furthermore, these guides refer to the social and cultural differences that lead an agent to decide (Awad et al. 2019). In machine ethics, the work of Anderson and Anderson is quite notable. In addition to their many research articles, it is particularly illustrating their edited book *Machine Ethics* (2011). For them, “it is better for machine ethics to be principle-

based" (Hooker and Kim, 2018-130 following Guarini, 2011), which implies non-consequentialist elements to respect dignity.

In the context of business ethics, companies carried out ethical codes of AI to address the problem of self-regulation (Vidgen, 2020). However, from other areas, the need for legislation to clarify the horizon and establish limits for these tools has been investigated. Thus, digital governance appears as the new challenge of technological innovation (Floridi 2018; Floridi and Taddeo 2016; Kaplan and Haenlein 2020), in which ethics must be considered both in the drafting of legislation as at the time of ensuring legal compliance. Likewise, for governance to be effective, it is necessary to focus not only on the implementation of innovations but also to attend to the entire process: from the design, development, deployment, and audit of AI systems (Kroll 2018).

To analyze AI ethical guidelines proposed in academic literature, we conducted a systematic literature review. We examined all the references indexed in the Web of Sciences (WoS) up to January 2020. We collected the data using the keywords "artificial intelligence" and "ethics," we found 1,370 documents; and refined the search to "journal articles" and to areas of social sciences and technology, which gave us a total of 262 study units, and 13,415 references cited in them. By selecting the research areas, we wanted to make sure we get everything published on AI ethics in the three main disciplines of the study: then we searched in management, philosophy, and technology journals. We were not interested in journals from areas such as medicine or other health sciences.

Our study showed us that there is a lack of definition of *the role of stakeholders*. According to the information analyzed, we identified several profiles in AI and business ethics scenarios like companies, employees, citizens (which are users of platforms and providers of data), governments, and practitioners (like engineers). It is essential to

delineate each of the parties' roles and responsibilities to enhance responsible innovation of AI for business.

To undertake the task of defining the role of stakeholders, we have focused on the designers' role for management decision-making (and developers) and on the role of companies that use algorithms in their decision-making process (here framed as users). There we analyze the ethical implication of AI decision-making processes from the perspective of ethics of care. According to Weltzien Hoivik and Domenec Melé (2009), "within business organizations, ethics of care focuses on relations between persons, on such relations as trust, mutual responsiveness, and shared consideration."

Since AI is usually associated with negative connotations for society and usually related to unemployment, lack of privacy for citizens, or algorithms' bias discrimination; We see the study of the role of AI from the perspective of care ethics as the perfect *antidote* to ameliorate some of the essential AI issues in business and organizations. Especially, the importance of social relations and context that the notion of care brings. AI is here to stay, and we must make the best out of it.

Today, there are three major ethical theories aplied in business ethics: deontology, utilitarianism, and virtue ethics (Melé 2014). Of these three, utilitarianism is the most used to analyze phenomena in the business world (Cranenburgh and Arenas 2014; Sandin 2016). However, the ethics of care is a relatively novel moral theory that would serve to identify and analyze the responsibility of firms with what regards especially to the most marginalized stakeholders. In this chapter, we will also take ethics of care as a complementary aspect of justice (Cavanagh et al. 1995), not limited to the personal, but connected to society.

4.3. THE ETHICS OF CARE AND BUSINESS ETHICS

The ethics of care appeared as a theory with Carol Gilligan's and Nel Noddings' approaches to caring. In her book *In a different voice* (1982), Gilligan presented care as a psychological theory for women's development, which is why it is often related as a feminist ethic (Borgerson 2007). Nevertheless, as Seigfried (1989, cited in French and Wis 2000) points out, a *different voice* is not limited to women but extends to men and is influenced by different social, political, and economic contexts.

The ethics of care appears as a response to the orthodoxy of ethics of justice since it is not bolstered on inviolable impartial principles but instead appeals to care relationships for personal well-being (French and Weis 2000; see also Held 2006). The perspective of care put aside the general standard to ask about the concrete situation and give an answer concerning circumstances and context (Gilligan 1982); This implies a moral vision centered on the individual. As Weltzien and Melé (2009) explained, while justice responds to ethical principles and duties, the theory of care focuses on attention to people's needs and their relationships; in this scheme, care is taken as a fundamental category, understood as a value and as an activity. In this sense, it can be argued that the ethics of care proposes solutions according to the interests of each party and not to previously established norms (Reiter 1996).

Since its inception, the notion of care has been developing, starting from the first definitions of care that seemed more ambiguous into a more rigorous definition. Based on previous works (as the work of Bubeck 1995; Clement 1996; Engster 2007; Fineman 2004; Held 2006; Kittay 1998; Noddings 2002; Slote 2001, 2007; Tronto 1993; Walker 1998; and With 2000, cited in Engster 2011), Daniel Engster (2011-98) proposed a definition of care ethics 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." That is the definition that we will follow.

The ethics of care have been applied to different areas in the business world since the 1990s (Melé 2014). Also, there is an insightful edited handbook about the topic, *Applying Care Ethics to Business* (Sander-Staudt and Hamington 2011). This theory has been used to analyze the pro-environmental behavior of employees (Paillé et al. 2016), also to investigate corporate philanthropy (Cranenburgh and Arenas 2014), crisis management (Sandin 2009), as well as when analyzing labor relations in small companies at a local level (Lähdesmäki 2019). However, this theory has not yet been used to study the ethical dimension of technological innovations in business organizations or analyze AI's role and its ethical implication in management decision-making.

Our theoretical proposal is adequate to clarify stakeholders' responsibilities to satisfy social trust in AI tools (Burton and Schoville 1996). With this, we entered into a research stream that avoids a contractualist vision of ethics. The last implies that the perspective of care eludes business codes of conduct and utilitarian policies in which the greatest amount of good is sought for the greatest number but is based on reciprocity (Lähdesmäki et al. 2019). This does not mean that other ethical views are not necessary or that they are not adequate but determine that we want to make a theoretical contribution to AI ethics in business bolstered on an approach emphasizing the person (Melé 2009). According to Sander-Staudt and Hamington (2011), by adopting this theoretical approach, we conceptualize mutual interdependence and cooperative relationships as ontologically essential.

4.3.1. Care and the Stakeholder Theory

Stakeholder theory is one of the most prominent theories in business ethics, R. Edward Freeman's work has been essential in its placement. In his pioneering publication, *Strategic Management – A stakeholder Approach*, Freeman (1984, p.46) gave the following definition: "A stakeholder in an organization is (by definition) any group or individual who can affect or is affected by the achievement of the organization's objectives." Since then, the stakeholder concept has developed in the discipline. In fact, since his publication in 1984 Freeman has given some specifications of the concept, some of the most important in his work with Gilbert in 1989 and 1992, and with Martin and Parmar in 2007. For our interests here, we will focus on the reinterpretation of the stakeholder concept in its approach from the point of view of the ethics of care. Specifically, we will take the article of Wicks, Gilbert, and Freeman, (1994), the work of Burton and Dunn (1996), and the discussion of this reinterpretation and proposal of Engster (2011).

In *A Feminist Reinterpretation of The Stakeholder Concept*, Wicks, Gilbert, and Freeman (1994) stated that a "feminist ethic" helps to "better express the meaning and purposes of corporations," and that the stakeholder concept with a feminist reinterpretation "yields important insights for corporation that want to improve their adaptability and responsiveness" (p. 477). The authors explained that behind the stakeholder concept are some masculine metaphors that shape business thoughts since the theory is for describing how business operates and for defining its basic purposes. The authors look at the following five specific metaphors:

> 1) the notion that corporations should be thought of primarily as an "autonomous" entity, bounded off from its external environment; 2) that corporations can and should enact or control their external environment;

> 3) that the language of competition and conflict best describes the character of managing a firm; 4) that the mode of thinking we employ in generating strategy should be "objective"; and 5) that corporations should structure power and authority within strict hierarchies.

These metaphors create the vocabulary and framework we use to understand the business world, the organization, and its purposes. Based on the propositions of care ethics, Wicks et al. proposed to; 1) see corporations as *webs of relations among stakeholders*; 2) to embrace change and uncertainty as dynamic and enriching forces for corporations; 3) to take communication and collective action as a form to resolve conflicts; 4) to not eliminate solidarity and empathy in business, but rather to use them as a strategy; 5) Finally, to *replace hierarchy with radical decentralization and empowerment.*

Following Wicks et al. (1994), Burton and Dunn (1996) stated that stakeholder theory as "a method of management based on morals and behavior" is missing a moral ground that traditional ethics cannot completely fulfill since it must recognize relationships among stakeholders. Then the authors propose "feminist ethics" (but referring accurately to the ethics of care) as moral grounding to provide the missing element of the stakeholder theory approach to management; and then they suggested a general principle for business decisions under the notion of care (primarily based on Gilligan's and Noddings' work). The principle state as follows: "Care enough for the least advantaged stakeholders that they not be harmed; insofar as they are not harmed, privilege those stakeholders with whom you have a close relationship." According to the principle, corporations must avoid harm in all decisions, and although it may not eliminate harm, the principle tries to limit it among the most vulnerable stakeholders.

Both propositions of a care-based stakeholder theory (the ones of Wick et al. and Burton and Dunn) took the principles of the ethics of care from the point of view of a then established "feminist ethics." Moreover, both works talked about "feminist ethics," not an "ethic of care," or "care ethics," although they referred to "care" as a central concept or mentioned a moral theory of care. The fact that the ethics of care ethcs was in its beginnings may have collaborated in this conception, but even then, Wicks et al. (1994) explained that "to speak of the 'care perspective' is not to speak only–or even primarily–to women, but to essential moral sentiments that we all share" (p. 478).

Bolstered on a developed definition of the ethics of care (and with the conviction of the insufficiency of Burton and Dunn principle to guide management decision), Daniel Engster (2011) proposed three care-based guidelines for the distribution of resources among stakeholders. The first distributional guideline is the *proximity principle*, which states that since our care resources are limited, we are justified to a) care for ourselves before others; b) to first care for individuals geographically and temporally close to us before others, and; c) to care for individuals in our own culture or state before others (based on Engster, 2007). The second guideline is the *relational principle*, which implies that a relationship's closeness relies on the dependency of one of the parts to meeting his or her needs. The third and last guideline is the *urgency principle* that proposes giving priority to those who need us to survive or function. According to Engster, these three guidelines should serve as priority rules for the distribution of care, since without this kind of principles and with the intention of everyone caring for everyone, care ethics would collapse (Engster 2011, p.98).

4.4. THE ETHICS OF CARE IN THE AI ERA: AN APPROACH FOR MANAGEMENT DECISION-MAKING

As discussed above, in our literature review we find a lack of definition of the role of the stakeholders in the AI era. The identification of the role of stakeholders is essential to delineate the responsibilities of each party. In this scheme, we studied the role of companies that adopt AI models and the accountability of corporations and engineers when using and designing algorithms for management decision-making.

Taking the work of Wicks et al. (1994) of a "feminist" (or care) approach to the stakeholder concept and the information analyzed in our literature review, we state that AI models, when applied to corporations, exacerbate the five masculine metaphors that shape our understanding of the business world, since they are behind the stakeholder concept, as explained by Wick, Gilbert and Freeman. Particularly, we see this in metaphor number 4: *that the mode of thinking we employ in generating strategy should be "objective."* Since one of the primary purposes of AI models and algorithms applied to management is to objectify decision making and to do it in an accelerated way (regardless of whether true objectivity is achieved). An example of this is that algorithms are used to determine who is offered or denied a mortgage, a loan, or who is hired or fired, all in the function of pre-established bias without considering the person and their circumstances. To alleviate this "stream" of a "masculine" interpretation of business in the AI era, we propose a care-based principle for the design of algorithms for management decision making.

Different authors have studied the accountability of algorithms and their ethical implications. In her work, Kristen Martin (2019a, 2019b) conceptualizes algorithms as value-laden and not as neutral (2019b) "in that algorithms create moral consequences, reinforce or undercut ethical principles, and enable or diminish stakeholder wight and

dignity" (p. 835). In this scenario, firms (as users of algorithms) and engineers (as developers) are responsible for the correct use of algorithms in management decision-making and for the error that may occur in the process (Martin, 2019a). What these arguments state is that in the first place if someone uses an algorithm to make a decision in a firm and then make a mistake, even if the mistake is unintentional, the firm would be responsible for ignoring or foster that mistake. In the second place, Martin stated that designers, while creating inscrutable algorithms, take accountability for their role in a decision. For example, in cases where companies that use algorithms have a minimal role in the decision and have no way to understand the procedure of the decision that is being made, the designer would be accountable (Martin, 2019b). Understanding the set of the two roles of firms and developers, in the use of algorithms in management decision making is essential to perform responsible decisions. So, this is why our principle is for both: for engineers designing algorithms and for corporations using them as tools when making decisions.

Following the principles of Burton and Dunn (1996) and Daniel Engster (2011), our principle would state as follow:

> For decision making in business, an algorithm should be designed to avoid harm to any stakeholder and insofar as no stakeholder is harmed, to distribute care according to the proximity, the relational or dependency, and the urgency of the needs of corporations' stakeholders.

In this sense, engineers are responsible for considering this principle in designing and developing the algorithm and firms in its application when making decisions. However, stated to the concrete use of developers the principle would be as follow:

> When developing an algorithm for decision making in business, designers should try to avoid the possibilities for it to be used to harm any

> stakeholder. Insofar as no stakeholder is harmed, the algorithm should, from its design, promote the distribution of care according to the proximity, the relational or dependency, and the urgency of the need of corporation's stakeholders.

Let us take as an example of the application of this sub-principle the fact that a developer could block in the design of his/her algorithm the possibility of a human resources manager to use variables such as ethnicity or socioeconomic level to decide whether to promote an employee. Or in the case of an algorithm for the finance sector and the decision to grant or not a loan, from its design the algorithm must avoid the possibility to relegate or marginalized certain sectors of society, or classes, to which loans are never granted and therefore they cannot advance in the purchase of family houses or in the founding of companies or other entrepreneurial projects.

The other side of the principle would be an application to the concrete use of firms, in that line the sub-principle would be stated as follow:

> When buying and using an algorithm for decision making, firms should ensure that when applying the algorithm, the output that it proposes, avoid harm to any stakeholder and insofar as no stakeholder is harmed, to distribute care according to the proximity, the relational or dependency, and the urgency of the needs of corporations' stakeholders.

If we take again the examples presented before, in the case of a human resources manager, the firm should be aware that certain algorithms could use undesirable variables (as sex, ethnicity, and the like) to decide to hire or not, or to promote or not someone. That could be the same case in the finance sector. It is in the responsibility of developers to avoid harm from the design of AI, and it is in the accountability of firms to buy and utilize AI that avoid harm to all its stakeholders. Although the design is in the capacities

of the developers, firms should be aware that accountability implications and to respond to their stakeholders would be its task. Since it is the company that will be really affected if it misuses AI tools.

In its application, this principle (and its sub-principles) should be supported for the proposal of adding social embeddedness and reflection in algorithmic decision-making process (Martin, 2019a). According to Kristen Martin, the said implies that when managing based on algorithms, decisions should not be seen as inevitable, and the context should be acknowledged. Moreover, it is essential that reflection stays as a fundamental part of the decision-making process; when managing based on algorithms, "users do not question changes for the future, as if the algorithm and the surrounding decision-making assemblage offer the best we have to offer without mistakes" (Martin, 2019a, p.136). Then, what we aim to contribute to our principle is a care-based way to add social embeddedness and reflection to management decision-making process in the AI era.

4.5. CONCLUSION

AI appears as one of the most prominent tools for the upcoming years, and its ethical implications as one of the most critical challenges of its applications. The European Union, the OECD, and the G20 have adopted principles for the use of AI (European Comission, 2020; G20, 2019; OECD, 2019); these institutions affirm that they intended to create human center principles. However, they do not specify the implications of this idea of human being that serve as the basis of their principles. Care ethics appears as the right ground of these principles since, as we defended, it eludes utilitarian policies and focuses on the person and her circumstances. Our intention in this chapter was to contribute to the goal of an AI human-centered by opening the study of a care-based AI

within business ethics. The said means that we wanted to propose the theory of care ethics to bolster some of the major business ethics problems.

Our study of the literature in AI and ethics showed us a lack of the definition of stakeholders' roles; this fact can affect the good work of corporations with the arrival of AI and has new ethical implications in companies. Then, in this line, our chapter's specific aim was to propose a care-base principle for management decision-making in the era of AI, bolstered by the stakeholder theory. Our principle should direct engineers' work when designing and developing AI for decision-making; and managers, when using algorithms to make decisions.

Future research should study other stakeholders' roles in the use of AI in management decision-making based on care ethics, as the role of governments and citizens, and as those who can be affected by the decision of algorithms, i.e. when they are denied a loan or a mortgage[54].

[54] This is the first chapter I wrote, together with my advisor Professor Fernández-Fernández. The fact that we propose a principle within the ethics of care could lead to a misunderstanding and misrepresent the purposes of this dissertation since in this doctoral thesis I defend principles are not enough to ensure an ethical AI within business. However, as Ned Noddings defend: "Care theorists, like virtue ethicists, put limited faith in principles. We not disdain principles. We recognize that principles—for example, an injunction against lying—help to keep daily life running smoothly. We learn the rule and commit ourselves to it, and for the most part we would not consider breaking it. However, when a real conflict arises, the principle is of little help. We have to dig behind the principle to see what deeper value has engendered it. (N. Noddings, *Caring: A relational approach to ethics and moral education* (University of California Press, Berkeley, CA, 2013))". Hence, even though we propose a principle in this chapter, the general contribution of this dissertation is a call to go beyond principles and proposes a focus on some notions of the ethics of care that are essential to some ethical issues of AI: just as references to context and circumstances, a relational approach to vulnerabilities and interdependent relationships, and *voice* or to consider the "other" needs.

REFERENCES

Anderson, M. and Anderson, S. (2011). *Machine Ethics*, Cambridge University Press, New York, NY.

Anderson, M., Anderson, S. L., and Berenz, V. 2019. A value-driven eldercare robot: Virtual and physical instantiations of a case supported principle-based behavior paradigm. *Proceedings of the IEEE, 107*(3): 526-540. doi:10.1109/JPROC.2018.2840045

Aronowitz, S., and DiFazio, W. 1996. High technology and work tomorrow. *Annals of the American Academy of Political and Social Science, 544*: 52-67. doi:10.1177/0002716296544001005

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., and Rahwan, I. 2018. The moral machine experiment. *Nature*, *563*(7729), 59. doi:10.1038/s41586-018-0637-6

Bell, D. (1973). The Coming of Post-Industrial Society: A Venture in Social Forecasting. New York: Basic Books.

Borgerson, J. 2007. On the harmony of feminist ethics and business ethics. *Business and Society Review, 112*(4): 477-509.

Buchholz, R. A. and Rosenthal, S.B. 2002. Technology and Business: Rethinking and moral dilemma. Journal of Business Ethics, 41(1/2): 45-50.

Burton, B. K., and Schoville, J. G. 1996. Feminist ethics as moral grounding for stakeholder theory.*6*(2): 133-147.

Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., and Floridi, L. 2018. Artificial intelligence and the 'good society': The US, EU, and UK approach. *Science and Engineering Ethics, 24*(2): 505-528. doi:10.1007/s11948-017-9901-7

Cervantes, J., Rodriguez, L., Lopez, S., Ramos, F., and Robles, F. 2016. Autonomous agents and ethical decision-making. *Cognitive Computation, 8*(2): 278-296. doi:10.1007/s12559-015-9362-8

Coeckelbergh, M. 2013. Drones, information technology, and distance: Mapping the moral epistemology of remote fighting. *Ethics and Information Technology, 15*(2): 87-98. doi:10.1007/s10676-013-9313-6

Davies, P. W. 2002. Technology and Business Ethics Theory. *Business Ethics: A European Review*, 6(2): 76-80.

Dodel, M., & Mesch, G. S. 2020. Perceptions about the impact of automation in the workplace. *Information Communication & Society*, doi:10.1080/1369118X.2020.1716043

Engster, D. 2011. Care ethics and stakeholder theory, in Hamington, M. & Sander-Staudt, M. (Eds.). *Applying care ethics to business*. New York: Springer: 93-110.

Engster, D. 2007. The heart of justice: Care ethics and political theory. Oxford: Oxford University Press.

European Comissión. 2020. White Paper. On Artificial Intelligence – A European approach to excellence and trust. Retrieved from: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf (May, 2020).

Freeman, R. E. 1984. *Strategic Management: A stakeholder approach*. Boston, MA: Pitman.

Freeman, R. E. and Gilbert, Daniel R. 1989. *Corporate Strategy and The Search for Ethics*. Englewood Cliffs, NJ: Pretice-Hall.

Freeman, R. E. and Gilbert, Daniel R. 1992. Business Ethics and Society: A critical Agenda. *Business & Society*, 31: 9-17.

Freeman, R. E., Martin, K., Parmar, B. L. 2007. Stakeholder capitalism. *Journal of Business Ethics* 74,4: 303-314.

Floridi, L. 2018. Soft ethics, the governance of the digital and the general data protection regulation. *Philosophical Transactions of the Royal Society A-Mathematical Physical and Engineering Sciences*, *376*(2133), 20180081. doi:10.1098/rsta.2018.0081

Floridi, L., Cowls, J., King, T. C., and Taddeo, M. 2020. How to design AI for social good: Seven essential factors. *Science and Engineering Ethics*, doi:10.1007/s11948-020-00213-5

Floridi, L., and Taddeo, M. 2016. What is data ethics? *Philosophical Transactions of the Royal Society A-Mathematical Physical and Engineering Sciences*, *374*(2083), 20160360. doi:10.1098/rsta.2016.0360

French, W., and Weis, A. 2000. An ethics of care or an ethics of justice. *Journal of Business Ethics, 27*(1-2): 125-136. doi:10.1023/A:1006466520477

Ghobakhloo, M. 2020. Industry 4.0, digitization, and opportunities for sustainability. *Journal of Cleaner Production, 252*, 119869. doi:10.1016/j.jclepro.2019.119869

Gilligan, C. 1982. *In a different voice*. Cambridge, MA: Harvard University Press.

G20. 2019. G20 Ministerial Statement on Trade and Digital Economy. Retrieved from: https://www.mofa.go.jp/files/000486596.pdf (May, 2020).

Hamington, M. and Sander-Staudt, M. 2011. *Applying care ethics to business*. New York: Springer.

Held, V. 2006. *The ethics of care: Personal, political, global*. Oxford: Oxford University Press.

Hill, C., and Rothaermel, F. T. 2003. The performance of incumbent firms in the face of radical technological innovation. *Academy of Management Review, 28*(2): 257-274.

Hooker, J. N. and Kim T.W. 2018. Toward Non-Intuition-Based Machine and Artificial Intelligence Ethics: A Deontological Approach Based on Modal Logic, AIES' 18, February 2-3, New Orleans, LA, USA. (Conference Paper).

Huber, C., and Munro, I. 2014. "Moral distance" in organizations: An inquiry into ethical violence in the works of kafka. *Journal of Business Ethics*, *124*(2): 259-269. doi:10.1007/s10551-013-1865-1

Jonas, H. 1984. *The imperative of responsibility*. Chicago: University of Chicago Press.

Jonas, H. 1985. *Technik, medizin und ethik — zur praxis des prinzips verantwortung —*. Frankfurt: Suhrkamp.

Kaplan, A., and Haenlein, M. 2020. Rulers of the world, unite! the challenges and opportunities of artificial intelligence. *Business Horizons*, *63*(1): 37-50. doi:10.1016/j.bushor.2019.09.003

Kim, T. W., and Mejia, S. 2019. From artificial intelligence to artificial wisdom: What socrates teaches us. *Computer, 52*(10): 70-74. doi:10.1109/MC.2019.2929723

Kim, T. W., and Scheller-Wolf, A. 2019. Technological unemployment, meaning in life, purpose of business, and the future of stakeholders. *Journal of Business Ethics, 160*(2): 319-337. doi:10.1007/s10551-019-04205-9

Kroll, J. A. 2018. The fallacy of inscrutability. *Philosophical Transactions of the Royal Society A-Mathematical Physical and Engineering Sciences, 376*(2133), 20180084. doi:10.1098/rsta.2018.0084

Lahdesmaki, M., Siltaoja, M., and Spence, L. J. 2019. Stakeholder salience for small businesses: A social proximity perspective. *Journal of Business Ethics, 158*(2): 373-385. doi:10.1007/s10551-017-3707-z ER

Liker, J. K., Haddad, C. J., and Karlin, J. 1999. Perspectives on technology and work organization. *Annual Review of Sociology, 25*: 575-596. doi:10.1146/annurev.soc.25.1.575

Ma, Z. 2009. The status of contemporary business ethics research: Present and future. *Journal of Business Ethics, 90*: 255-265. doi:10.1007/s10551-010-0420-6

Martin, K. 2019a. Designing Ethical Algorithms, *MIS Quarterly Executive,* 18(2): 129-142.

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

Marx, K. 1932. The Economic and Philosophic manuscripts of 1844 and the communist manifesto. New York: Prometheus Books.

Melé, D. 2009. *Business ethics in action. seeking human excelence in organizations*. New York: Palgrave-MacMillan.

Melé, D. 2014. "Human quality treatment": Five organizational levels. *Journal of Business Ethics, 120*(4): 457-471. doi:10.1007/s10551-013-1999-1 ER

Misselhorn, C. 2020. Artificial systems with moral capacities? A research design and its implementation in a geriatric care system. *Artificial Intelligence, 278*, UNSP 103179. doi:10.1016/j.artint.2019.103179

OECD. 2019. OECD principles on AI. Retrieved from: https://www.oecd.org/going-digital/ai/principles/ (May, 2020)

Paille, P., Mejia-Morelos, J. H., Marche-Paille, A., Chen, C. C., and Chen, Y. 2016. Corporate greening, exchange process among co-workers, and ethics of care: An empirical study on the determinants of pro-environmental behaviors at coworkers-level. *Journal of Business Ethics, 136*(3): 655-673. doi:10.1007/s10551-015-2537-0 ER

Rachels, J. 2003. The Elements of Moral Philosophy. New York: The McGrw-Hill Companies, Inc.

Peace, A. G., Weber, J., Hartzel, K. S., Nightingale, J. 2002. Ethical Issues in eBusiness: A Proposal for Creating the eBusiness Principles. *Business and Society Review*, 107(1): 41-60.

Reiter, S. A. 1996. *The kohlberg–gilligan controversy: Lessons for accounting ethics education* doi:https://doi.org/10.1006/cpac.1996.0005

Remenyi, D., and Williams, B. 1996. Some aspects of ethics and research into the silicon brain. *International Journal of Information Management, 16*(6): 401-411. doi:10.1016/0268-4012(96)00029-1

Robbins, R. W., and Wallace, W. A. 2007. Decision support for ethical problem solving: A multi-agent approach. *Decision Support Systems, 43*(4): 1571-1587. doi:10.1016/j.dss.2006.03.003

Sander-Staudt, M., and Hamington, M. 2011. Introduction: Care ethics and business ethics. *Applying Care Ethics to Business, 34*, VII-+. doi:10.1007/978-90-481-9307-3

Sandin, P. 2009. Approaches to ethics for corporate crisis management. *Journal of Business Ethics, 87*(1): 109-116. doi:10.1007/s10551-008-9873-2 ER

Schwab, K. 2016. *The fourth industrial revolution*. Geneva: World Economic Forum.

Shank, D. B., Graves, C., Gott, A., Gamez, P., and Rodriguez, S. 2019. Feeling our way to machine minds: People's emotions when perceiving mind in artificial intelligence. *Computers in Human Behavior, 98*: 256-266. doi:10.1016/j.chb.2019.04.001

Van den Hoven, J., Vermaas, P. E., Van de Poel, I. (2015). *Handbook of ethics and values in technological design*. Dordrecht: Springer.

Van de Poel, I. (2018). Design for value change, *Ethics and information technology*. Retrieved from: https://doi.org/10.1007/s10676-018-9461-9.

Van Cranenburgh, K. C., and Arenas, D. 2014. Strategic and moral dilemmas of corporate philanthropy in developing countries: Heineken in sub-saharan africa. *Journal of Business Ethics, 122*(3), 523-536. doi:10.1007/s10551-013-1776-1 ER

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: 491-501.

Wiener, M., Saunders, C., and Marabelli, M. 2020. Big-data business models: A critical literature review and multiperspective research framework. *Journal of Information Technology, 35*(1): 66-91. doi:10.1177/0268396219896811

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.

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. doi:10.1016/j.bushor.2018.07.001

Yuthas, K. and Dillard, J.F. 1999. Ethical Development of Advanced Technology: A Postmodern Stakeholder Perspective. Journal of Business Ethics, 19(1): 35-49.

CHAPTER 5

ARTIFICIAL INTELLIGENCE AND CORPORATE RESPONSIBILITY

How and why firms are responsible for AI

In *Encyclopedia of Business and Professional Ethics*. Poff, D. and Michalos, C.M. Eds.

Springer Nature. ISBN-13: 978-3030227654

When a firm develops an AI program, that firm makes value-laden decisions as to who is important, who should be considered, and who can be ignored in a given decision. For example, in a mortgage approval program, the computer scientists train the algorithm on previous applicants including who was approved and rejected over a number of years. The AI program 'learns' the attributes of individuals who are more likely to be approved. In any given data set, some people will be well-represented with all the data filled out and some will not have all their data included. Some types of people will be completely missing from the data. Data and computer scientists need to decide how much to punish people who are not represented or not well represented in the data. In addition, these same data and computer scientists make assumptions about missing data, how to treat outliers or edge cases, and how morally important it is to include more people in the model. In other words, if the predictive mortgage approval model does work well with certain people, should we care? Does it matter? How much should a bank care?

All this is to say that the firms that develop AI programs make value-laden decisions during design and development (Martin 2019). And that these decisions have moral implications for the firms that adopt the AI program and the users who are subject to a particular AI program. This runs counter to the mistaken belief that AI is somehow neutral or operates outside human involvement. In fact, these data and computer scientists have to make value-laden decisions throughout the development process.

- Training and Live Data. When algorithms are developed from training data, who is represented in the training data and how the data is labeled directly impacts the creation of the algorithm. For example, when facial recognition is trained on

primarily white men, the result is an algorithm who identifies white men moderately well but identifies black women incorrectly the majority of the time (Buolamwini and Gebru 2018). The model that is developed on a specific training data set may also be tailored to that training data and ineffective when applied to live data, causing harms, breaking rules, and reinforcing existing power dynamics.

- Development of the Model. Computer scientists make assumptions about the type of data, how the data is distributed, whether data is missing (and how bad is it for data to be missing), whether the algorithm should care about outliers (and how much should it care). These are all value-laden decisions about individuals.
- Outcome Chosen. How does a particular outcome favor certain groups of people and how well does the outcome represent the phenomenon of interest? For example, we use GPA as a measurement for 'good student in college' sometimes, but that does not mean that the GPA as an outcome is a good measurement of the phenomena we are interested in.
- Mistakes. All AI programs generate mistakes – people are mischaracterized and misidentified. Sometimes AI predicts someone will commit a crime and they do not, which is a false positive. Other times AI programs will predict someone will not commit a crime and they do which is a false negative. The types of mistakes (false positives versus false negatives) vary across decision contexts as well as which mistakes are more preferable for a given decision. For example, in the criminal justice system, we *prefer* false negatives: we prefer in the United States that someone be falsely set free rather than falsely imprisoned. Not only do computer scientists influence that types of mistakes that are more common with a given AI program, but they also influence whether or not the inevitable mistakes are able to be identified, judged, and fixed by users of the AI program. AI

programs that are developed to be inscrutable, e.g., declared proprietary or designed to not be accessible by the firm that uses the AI program, allow the inevitable mistakes to continue by not allowing users to identify, judge, possibly fix mistakes.

- Contestability. While people like to think that AI and related computer programing approaches are inscrutable, computer scientist Joshua Kroll notes that "inscrutability is not a result of technical complexity but rather of power dynamics in the choice of how to use those tools" (Kroll 2018). In other words, making a program difficult to use or making the mistakes created by the program difficult to identify, judge, and correct is a design decision. In fact, developers of AI programs should make their programs contestable (Mulligan, Kluttz, and Kohli 2020), where subjects of the AI program are able to contest any decision made about them. This would require a certain amount of transparency and accountability in the design depending on the context of the decision and the types of users subjected to the program.
- Assessment of AI. The computer scientist influences how the AI program is assessed that it 'works.' While we regularly, in the popular press and in academia, claim that AI is 'accurate' or 'efficient,' these measurements are actually constructed in the design for many programs. For example, one might need to know for whom is the program accurate and for whom is it not accurate. And, the efficiency gains for a company implementing AI programs may also mean that a bad decision is being made faster. We normally do not see mere efficiency as a goal for decision making. If we are hiring or arresting the wrong people, making those types of decisions faster with the aid of AI does not make the entire organization more efficient and may offload some of the work onto others. In

fact, even the idea of prioritizing claims of accuracy and efficiency is a value-judgement that may work for the developing firm but not for the firm adopting and using the AI program (Johnson, forthcoming).

5.1. WHY FIRMS ARE RESPONSIBLE FOR AI

While the data and computer scientists make value-laden decisions in developing the AI program, the firm that *uses* the program is responsible for the ethical implications of their business decision. In other words, the bank is still responsible for making mortgage decisions, insurance companies are still responsible for adjudicating insurance claims, and firms are still responsible for their hiring decisions *even if* they augment their decision with an AI program. This places a distance between the moral decisions of development and the ethical implications in use.

Hence, the introduction of AI to decision-making increases what scholarship has called moral distance. Scholars use this concept to explain why individuals behave unethically towards those who are not seen. With AI decision-making, face-to-face interactions are minimized, and decisions are part of a more opaque process that humans do not always understand. Therefore, the issue regarding AI and moral distance is that firms miss the moral implications of their decision, for which they are responsible, being blinded behind the veil of AI (Villegas-Galaviz and Martin).

Firms are responsible for the development, deployment, and use of AI in the same manner these same firms are responsible for the many business decisions they make about the products they develop, the materials they purchase, and the decisions about individuals that they make. Firms are responsible for the products and services they sell in that they have an obligation to not cause harm, act in a manner that does not further disadvantage the less fortunate, abide by the values and norms of society, and follow the

law. Firms are similarly responsible for the decisions, augmented with AI, they make about individuals, employees, and users in that they have an obligation to treat people with dignity and respect, act as if individuals are an end and not a mere means to be used merely for the firms benefit, and to not create harm or diminish rights. The introduction of AI into an organization does not remove their responsibility for their actions.

5.2. APPROACHES TO TAKE RESPONSIBILITY FOR AI

Our ethical concepts, traditions, theories, and approaches can be seen as a way to close the gap between those making value-laden decisions and the ethical implications of those same decisions. In other words, these theories and approaches help the data and computer scientists understand better the ethical implications of their work. And, for firms adopting AI, these approaches provide a roadmap of the types of questions one should ask about the design and development and use of a specific AI program. Here we focus on more than mere consequentialism, which would only ask firms to calculate the possible net benefits or harms caused by the development, deployment, and use of AI. Consequentialism has the same deficits as an ethical tool when applied to AI decisions: the harms to the few who are considered marginalized, without a voice, or 'edge' cases can be ignored in order to benefit the more powerful. Instead, we focus on those ethical approaches that would help firms take responsibility for AI and decrease the moral distance exacerbated by the use of AI.

5.2.1 Deontology

In the field of AI and business ethics, much work has been done to find the right set of principles or AI ethical guidelines. Deontology, or principle-based ethics, bases the rightness of the action in that it follows the duty of those who act. Hence, individuals should decide according to their principles or rules rather than considering the consequences or context. These attempts within AI Ethics usually follow the line to bring ethical frameworks from other disciplines, especially the four essential principles traditionally used in bioethics: beneficence, nonmaleficence, autonomy, and justice. However, scholars have brought out the fact that principles are not sufficient to guarantee ethical AI and the limitations of a principled approach to AI ethics (Mittelstadt 2019).

5.2.2. Justice and Fairness

Fairness and AI has become almost synonymous with ethical AI, primarily when AI has been used to sort individuals, the program reinforces existing injustices captured in the data. For fairness and justice approaches to AI, initial works focused on how algorithmic decision-making processes do not lead to more objective and or more fair decisions than those by humans, who are seen as influenced by prejudice or emotions. In fact, AI has the potential to exacerbate issues regarding discrimination, bias, and fairness. In applying fairness approaches, best practice is to distinguish questions about discrimination and questions about justice. Fairness and justice theories highlight how being predicted or categorized should not be more likely for particular groups of people and that the system of allocating goods (the AI program) should not harm the less fortunate. Other approaches focus on equity, parity merit, and even the appropriateness of using particular attributes of individuals for a decision (Martin

2019). Discrimination law, on the other hand, focuses on ensuring the individuals are not treated or impacted differently based solely on a protected attribute (nationality, race, ethnicity, sexual orientation, gender, religion, etc), and problems of discrimination are best examined throughout the process of design, development, and use of AI (Barocas and Selbst 2016).

5.2.3. Virtue Ethics

Within virtue ethics approaches to understanding AI, the character traits of the agent or subject/user is the focus. Shannon Vallor's proposals lead the way to bring virtue ethics to answer the critical ethical questions of the current era. In her book *Technology and the Virtues* (2016), Vallor proposed a virtue-driven approach to the ethics of emerging technologies, such as AI, and a kind of ethical strategy for promoting the moral character needed for the challenges of recent times. In her framework, she adapted Aristotelian, Confucian, and Buddhist ethical reflections to create a set of what she calls are the *technomoral virtues* needed for the 21st century. The *technomoral virtues* framework 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. There, in search of a *technomoral wisdom* the framework proposes an adaptation of twelve virtues to the new techno-social environment, in there are virtues like honesty, self-control, humility or civility.

5.2.4. Ethics of Care

More recently, the ethics of care has been used to better understand the moral implications of AI. The ethics of care is a contextualized moral theory focuses on interdependent relationships, individuals' vulnerabilities, circumstances, and the voice of the other in ethical decision-making. In AI ethics, the contribution of the ethics of care comes in line with the understanding of how AI models may marginalize those who do not fit within the pattern created and used by those who develop and deploy AI. In its critical aspects, the ethics of care can help in the comprehension of how algorithm decision-making can create harm and ignore the needs of individuals, especially the most marginalized groups (Villegas-Galaviz 2022).

5.2.5. Critical Approaches

Critical theories attempt to understand the power dynamics and seeks to question not only the presumed objectivity and neutrality of analytics (Johnson, n.d.) but also the power dynamics at play in building the algorithm, collecting and using the data, and deploying AI and analytics. Critical approaches seek to understand who gains and who is marginalized by the status quo. Langdon Winner (1980) is perhaps the most well-known scholar to take this approach to technology more broadly. Winner argues that technology, designed and used by society, has politics or "arrangements of power and authority in human associations." In regards to AI, critical approaches examine the development and use of AI through the lens of power – who retains power and who is marginalized – and usually makes the case for the lifting or emancipation of those who are being undermined by the use of AI.

5.3. CONCLUSION

When considering the ethical implications of development or use of AI to augment decisions, business practitioners and business ethics scholars have the tools to better understand how AI can be developed and used within the given values of the firm. AI does not fundamentally change how we think about ethics and responsibility.

REFERENCES

Barocas, S., and Selbst, A.D. (2016). Big Data's Disparate Impact. California Law Review 104.

Buolamwini, J. and Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of Machine Learning Research, Conference on Fairness, Accountability, and Transparency, PMLR 81 (pp. 1–1)5. New York City, USA.

Johnson, G. n.d. Are Algorithms Value-Free? Feminist Theoretical Virtues in Machine Learning. Journal Moral Philosophy.

Kroll, J. A. (2018). The Fallacy of Inscrutability. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Science, 376 (2133): 20180084.

Martin, K. (2019). Ethical Implications and Accountability of Algorithms. Journal of Business Ethics 160 (4), 835–850.

Mittelstadt, B. (2019). Principles Alone Cannot Guarantee Ethical AI. Nature Machine Intelligence 1(11), 501–507.

Mulligan, D. K., Kluttz, D. and Kohli, N. (2020). Shaping Our Tools: Contestability as a Means to Promote Responsible Algorithmic Decision Making in the Professions. In Werbach, K. (ed.), After the Digital Tornado. Cambridge University Press.

Vallor, S. (2016). Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting. Oxford University Press.

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

Villegas-Galaviz, C., and Martin, K. Moral Distance, AI, and the Ethics of Care.

Winner, L. (1980). Do Artifacts Have Politics? Daedalus 109(1), 121–136.

CHAPTER 6

THE ETHICS OF CARE AS MORAL GROUNDING FOR AI

In Martin, K. (Ed.) *Ethics of Data and Analytics*. Taylor & Francis.

ISBN-13: 978-1032062938

> In information societies, operations, decisions, and choices previously left to humans are increasingly delegated to algorithms, which may advise, if not decide, about how data should be interpreted and what actions should be taken as a result. Examples abound. Profiling and classification algorithms determine how individuals and groups are shaped and managed (Mittelstadt et al., 2016).

Technology has always appeared as a way to expand human capacities (Jonas, 1979). Take, for example, the case of force and the augmentation of physical labor with the steam engine, or the case of the transportation sector and how it has been developed from long walks to wagons, then to bicycles, trains, cars, and airplanes, now ending up with autonomous vehicles. Thanks to technology, the whole conception of moving from one place to another has changed. Developments of technology have modified the character of human action. Taking this argument, the philosopher Hans Jonas (1979) argues that ethics has to do with actions and when changing the nature of human actions, there must necessarily be a kind of adaptation in ethics to new scenarios: to look to new approaches, ethical frameworks, and to ask new questions.

Today, emerging technologies such as AI are transforming the way that humans behave in society. “Algorithms silently structure our lives. Algorithms can determine whether someone is hired, promoted, offered a loan, or provided housing as well as determine which political ads and news articles consumer see” (Martin, 2019b). Previous technologies enhance human physical capacities, as in manufacturing, and others facilitate the storage and management of information, as in digitization, now AI is

changing how humans make decisions, and that impacts how humans' decisions affect others.

The delegation of human autonomy to algorithm decision-making has been studied from different fields (from law, from engineers, and philosophers), and some ethical principles and guidelines have been proposed by scholars (Floridi et al., 2018; Floridi et al., 2020, to name a few), governments (European Comission, 2020; G20, 2019; OECD, 2019), and enterprises (IBM, 2021). However, there is still so much to do within AI ethics and to ask and answer about responsibility, fairness, egalitarianism, values, and virtues in the AI era.

In this chapter, I focus on how the reduction of decision-making to data analytics may lead to moral dilemmas in how we make decisions about people: who is included and who is excluded. I will propose a care-based approach to shed light on how relationships, interdependence, vulnerabilities, and emotions should not be ignored.
Ethics of care appeared as a theory with Carol Gilligan in her book *In a different voice* (1982), where the author presented care as a response to the orthodoxy of ethics of justice. With the notion of care, Gilligan brought out the key argument regarding how relationships, interdependence, circumstances, and emotions are an essential part of ethical decision-making. The said imply that a reduction to formal rationality and an indifferent weighing of principles and norms is not enough in ethical terms.

In what follows, I briefly introduce some important facts of how AI works, then I present ethics of care to mitigate the moral problems presented in AI and decision-making. After that, I propose some questions that may serve as guidelines when applying AI while considering the notion of care.

6.1. TO FIT WITHIN THE PATTERN

AI is “defined as a system’s ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation” (Kaplan & Haenlein, 2019). Hence, using data as raw material, AI decision-making works by creating patterns and making predictions (Martin, 2019a). When using AI to decide on a particular individual in a specific circumstance, the result would be to judge that person as fitting or not fitting in a previously established pattern, and that pattern was created with previously gathered data. This means that the “fit” of that individual in a pattern is what determines the decision. Barocas and Selbst (2016) allude to this sequence of steps when explaining that data mining is a form of statistical discrimination where the use of AI reproduces past prejudices by identifying a pattern in the training data and then enforcing that pattern on new data (p. 675).

Decision-making with AI is done usually with the goal of maximizing efficiency, making decisions faster and supposedly more objectively. However, any efficacy enjoyed is for those who design and deploy AI. The AI decision is within their power, so in case of doubt, the resolution goes in their service. For example, AI helps to know more quickly and more "safely" who to hire or who not for a job, to whom to grant or deny a loan or mortgage, or to whom to grant it or not parole, as in the case of the COMPAS algorithm, used in court to grant or not parole. Hence, those who apply the model may be reducing that decision to data and ignoring vulnerabilities and specific circumstances that could be essential to decide morally. Not only does the use of AI codify patterns of the past, the application of that codified past ignores the vulnerabilities and specific circumstances of the subjects present. That is why AI may affect the most marginalized stakeholders, and why big data processes could improperly disregard legally protected classes, leading to a *disparate impact* that “refers to policies or practices that are facially neutral but have a

disproportionately adverse impact on protected classes" (Barocas and Selbst, 2016, p. 694).

Making decisions using AI is about excluding those that do not fit a pattern and including those that fit within the pattern. And many have examined those that do not fit, who are marginalized or left behind or discriminated against with AI programs with the lens of justice (Mittelstadt et al., 2016; Coeckelbergh, 2020; Dubber et al., 2020). However, in the study of the ethical challenge of *those that fit and do not fit,* those who are elevated and those who are marginalized, I propose the theory of ethics of care as moral grounding for the AI era. In what follows, I am going to briefly explain care ethics and then propose it as a way to alleviate the moral problems presented.

6.2. ETHICS OF CARE

In her book *In a different voice* (1982), Carol Gilligan presented care as a response to the orthodoxy of ethics of justice. Gilligan first presented care as a psychological theory for woman's development. However, with the notion of care, the author brought out the key argument of how relationships, interdependence, circumstances, and emotions are essential parts of ethical decision-making. That means that focusing solely on formal rationality and principles is not enough for morality.

> "This conception 'of morality as concerned with the activity of care centers 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)

For Gilligan, the general idea of care is to understand responsibility and morality in the context of relationships and to resolve moral dilemmas in the comprehension of

dependence and vulnerability. Communication plays an essential role since it is the way to listen to relational voices and listen to a different voice. Where "to have a voice is to be human. To have something to say is to be a person. But speaking depends on listening and being heard; it is an intensely relational act" (Gilligan, 1982). Therefore, care should be taken as a *practice and a work that must be done on a direct level* (Sander-Staudt & Hamington, 2011). In this sense, the perspective of care implies to decided considering the person in her specific circumstances and not based on previously established norms (Reiter, 1996).

Since coined, the notion of care has developed to a more rigorous definition of the term, now not only linked to woman's development (French and Weis, 2000). Based on previous literature on care, Daniel Engster (2011, p. 98) proposed a definition of care ethics 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." This definition encompasses what implies going beyond the formal rationality based on principles, guidelines, and norms within ethics.

6.3. A CARE-BASED AI

As the philosopher Hans Jonas defended, ethics must adjust to the changes that technology produces while expanding, increasing, and transforming the nature of human actions. The said adaptation suggests a look to new frameworks, a reconsideration of how theories are applied, and an invitation to ask new ethical questions.

AI is being used to categorize people, to elevate those who fit and marginalize those who do not fit a particular pattern. The artist Mimi Onuoha defined *algorithmic violence* as "the violence that an algorithm or automated decision-making system inflicts

by preventing people from meeting their basic needs."55 Those that are marginalized or do not fit a particular pattern are then denied rights or further harmed feel that algorithmic violence. Hence, a theory that put vulnerability, harm, and relationships in the foreground would better identify wrongs of AI decision making.

The theory of ethics of care would help to better understand the moral implications of algorithms. Based on previous research on the ethics of care, we can preset the following five categories as key elements to understand a care-based ethics of AI decision-making. Each category is presented with some questions that those who develop and deploy AI should have in mind when applying ethics of care to data analytics.

6.3.1. Interdependent Relationships

Within the ethics of care, responsibility and morality have a meaning in a web of interdependent relationships. That means that "the admonition to maintain relationships, and to be cognizant and responsive to the needs of others, are two general principles central to an ethic of care. Nevertheless, more than providing such principles, an ethics of care recommends itself as a method and way of orientating oneself towards the world" (Sander-Staudt & Hamington, 2011). To understand accountability in a network of relationships means to put aside the general standard and to look to concrete situations, where "the generalized other" becomes "the particular other," a specific individual in a particular circumstance (Gilligan, 1982).

When applying ethics of care to AI, it would be essential that models do 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). There we would ask:

55 Retrieved from: https://mimionuoha.com/algorithmic-violence (July, 2021).

- Which interdependence relationships can be affected by the development of this algorithm?
- Are relevant interdependence relationships being ignored in the development of this model?
- Would emotions be an eliminated essential part of the kind of decision that is being automated?

6.3.2. Context and Circumstances

According to care ethicists, care is a practice and something to be done on a direct level, a face-to-face interaction. Also, care may be understood as a "motive, ideal, virtue, and method." However, "care" should be distinguished from "personal service", "the former involves meeting the needs of those who are unable to meet such needs themselves, the latter involves meeting needs for others who could meet such needs themselves." (Sander-Staudt & Hamington, 2011; see also Bubeck, 1985).

For AI, the said imply that those affected by AI decision-making will not have the possibility of meeting their needs. They depend on the algorithm to do it. Hence, there is a moral responsibility to care for them. Moreover, for AI ethics, the relevance of the "direct level" involves a more contextual mode of judgment and the awareness that decisions should not result from an abstraction of the problem, eliminating the context. Having this in mind, one should ask:

- Does this algorithm imply the elimination of context and circumstances when they can be an essential part of a future decision?
- Does this model open the possibility to social and cultural embeddedness?

6.3.3. Vulnerability

The notion of care implies the comprehension of the vulnerability, the needs, and suffering of the other. Also, in a network of interdependent relationships, everyone becomes vulnerable, and there appears care as an essential concept. "When we care for individuals, we usually aim to help them to meet their basic needs, develop or maintain their basic capabilities, or alleviate their pain and suffering." (Engster, 2011). That means that care includes all that is in line to meet everyone's basic needs.

Care-based AI implies that algorithms do not prevent individuals from meeting their needs, especially the most basic ones. When applying ethics of care to AI, we would ask:

- Does this algorithm prevents the possibility of fostering the needs of protected classes, people at risk of social exclusion, or marginalized stakeholders?
- Does the data used imply exploiting the vulnerabilities of those affected by this algorithm?
- Are vulnerabilities used as variables to prevent future enhancement for those affected by this algorithm?

6.3.4. Voice

As presented by Carol Gilligan (1982), voice means to have the possibility of defending one's own interpretation and needs. For example, Gilligan says that "to have a voice is to be human. To have something to say is to be a person. But speaking depends on listening and being heard; it is an intensely relational act." (Gilligan, 1982). That means that it is essential to give voice to every affected part in a situation with ethical implications. Also, voices should be heard through communication in relationships.

There, communication is presented as the method of conflict resolution and the way to resolve moral dilemmas because it gives the possibility to hear different voices.

For AI, this would mean that algorithm should maintain open the possibility of hearing different voices and not silent any voice that should have part of the situation in which it is applied. For this purpose, interdisciplinary teams could serve to comprehend the different points of view to try to hear the voices of different cultures and social collectives. Hence, having this in mind, we would ask:

- Which voices are being silenced with the development of this algorithm?
- Furthermore, does this algorithm hear all the different voices needed? For example, in an interdisciplinary way?

6.4. CONCLUSION

The purpose of this chapter was to bring out how the reduction of decision-making to data analytics may lead to a moral problem where people's opportunities are reduced to their fit into a previously created pattern. In this context, ethical theories as deontological ethics, utilitarianism, consequentialism, and ethics of justice are necessary but not sufficient. There I proposed ethics of care as moral grounding for the AI era. The notion of care appears as an essential key to alleviate the moral problems derived from a tendency to look for apparent objectivity bolstered in efficiency for decisions. Ethics of care may serve to shed light on the fact that considering vulnerabilities and interdependence relationships is fundamental to morality. Some essential notions of ethics of care were provided to serve as key elements to understanding the ethics of AI decision-making.

REFERENCES

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review, 104*(3), 671-732.

Bubeck, D. (1985). *Care, justice and gender*. Oxford: Clarendon Press.

Coeckelbergh, M. (2020). *AI Ethics*. Cambridge, MA: The MIT Press.

Dubber, M., Pasquale, F., Das, S. (2020). *Oxford Handbook of AI*. New York: Oxford University Press.

Engster, D., 2011, Care ethics and stakeholder theory, in Hamington, M. & Sander-Staudt, M. (Eds.). Applying care ethics to business. New York: Springer, pp. 93-110.

Engster, D. (2007). *The hear of justice: Care ethics and political theory*. Oxford: Oxford University Press.

European Comission. (2020). Ethics guidelines for trustworthy AI. Retrieved from: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (July, 2021).

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., . . . Vayena, E. (2018). AI4People-an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*(4), 689-707.

Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2020). How to design AI for social good: Seven essential factors. *Science and Engineering Ethics, 26*(3), 1771-1796.

French, W., and Weis, A. (2000). An ethics of care or an ethics of justice. *Journal of Business Ethics, 27*(1-2): 125-136.

G20. (2019). G20 Ministerial Statement on Trade and Digital Economy. Retrieved from: https://www.mofa.go.jp/files/000486596.pdf (July, 2021).

Gilligan, C. 1982. *In a different voice*. Cambridge, MA: Harvard University Press.

IBM. (2021). AI Ethics. Retrieved from: https://www.ibm.com/artificial-intelligence/ethics (July, 2021).

Jonas, H. (1979). *Das Prinzip Verantwortung.*

Kaplan, A., & Haenlein, M. (2019). Siri, siri, in my hand: Who's the fairest in the land? on the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons, 62*(1), 15-25.

Martin, K. (2019a). Designing ethical algorithms. *Mis Quarterly Executive, 18*(2), 129-142.

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

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3*(2), 1-21.

OECD. 2019. OECD principles on AI. Retrieved from: https://www.oecd.org/going-digital/ai/principles/ (July, 2021)

Sander-Staudt, M., and Hamington, M. (2011). Introduction: Care ethics and business ethics. *Applying Care Ethics to Business, 34*, VII.

Reiter, S. A. (1996). *The kohlberg–gilligan controversy: Lessons for accounting ethics education* doi:https://doi.org/10.1006/cpac.1996.0005

Conclusion

In general, the underlying objective of this dissertation was to analyze and respond, from an ethical perspective, to the question of *what we lose and what we put at stake when we delegate our decision processes to algorithms?* There to propose the ethics of care as a unique and much-needed approach to mitigate some ethical problems of the specific attributes of how AI works: blocking empathy in those who decide (caused by the distance between development and deployment of AI and its impact), reinforcing systems of power, unfairly rating and profiling individuals with data, and marginalizing vulnerable stakeholders.

Throughout this dissertation, emphasis has been placed on the fact that data and analytics are the leading forces of technology shaping business and society. Also, all these chapters explain how AI (there data and analytics) continues to be the source where firms and organizations search for responses and proposals, and the different ways in which the delegation of decision-making to AI can impact business and society.

Currently, there are many studies in the field of AI ethics, and several scholars are working on the specialized intersection between business ethics and AI. Also, in this dissertation, special attention has been placed on the ethics of care approach to business ethics and previous work on the ethics of care and technology. All those works have been explained. However, the contribution of this dissertation is the proposition of the ethics of care as an ethical framework to analyze AI within business.

The propositions of the ethics of care were framed and presented in this dissertation as categories to consider in the overall process of AI decision-making (from the design, development, and deployment to the use of AI in firms). These categories imply the relevance of considering 1) *interdependent relationships* in AI processes; the problem of the exploitation of 2) *vulnerability*; the role of 3) *context* and *circumstances* in decision-

making; and the significance of what the *other* has to say, to which we have referred as 4) *voice*.

At this point, it is critical to highlight three essential points that this dissertation did not propose but could be inferred in a wrong way:

- In general, the ethics of care has been proposed as a relevant approach to AI ethics. However, it is necessary to emphasize that the intention was never to present the theory as the only one needed. Just as presented in Chapter 1, the ethics of care emphasizes aspects sometimes ignored or overlooked by other moral approaches, features that are essential to AI ethics and business. However, the good development of AI ethics within business can only be achieved together with the rest of the approaches presented, such as virtue ethics, deontology, justice, and the like.
- Also, this compendium contributes to AI ethics with a critical examination of the technology. Hence, the research has focused on the possible harms of the technology. However, the main intention was never to argue for the elimination of AI as a tool or disregard the usefulness of algorithms to improve business decision-making. Instead, the proposal was to contribute to a better AI. The general idea is that if the categories proposed here are considered in AI processes, those could be improved.
- Finally, chapter three could be a summary of what this dissertation is not proposing: this compendium does not propose the development and deployment of AI from the point of view of altruism, partiality, nor even from a feminist approach that is only about women (but a relational one), and although it is aligned to virtue ethics, the proposition of this thesis focuses on different aspects than those of a virtue approach.

1. IMPLICATIONS FOR PRACTICE

This compendium explains how the impact of AI decision-making is responsibility of those who develop and deploy AI. The said imply that firms cannot justify the outcomes of their actions just by blaming the algorithm or stating that they do not really understand AI processes, moreover firms should comprehend the technology they are using. Hence, this dissertation aligns with the research stream that "conceptualize algorithms as value-laden, rather than neutral, in that algorithms create moral consequences, reinforce or undercut ethical principles, and enable or diminish stakeholder rights and dignity." (Martin, 2019a).

In Chapter 1, we explained how different moral frameworks focus on aspects that should be considered for practitioners in AI processes. In Chapter 2, we explained how firms and organizations should be aware of the problem of moral distance and how it can affect individuals when designing, developing, and deploying AI for business decision-making. Chapter 4 focused on the role of stakeholders and stated a care-based principle for those who develop and deploy AI, focused on avoiding harm. Chapter 5 offered an overview to practitioners to comprehend how and why firms are responsible for AI. Chapter 6 summarized why firms should be aware of the moral implication of profiling and classifying individuals with AI and how they can ameliorate the harm created by those processes. Finally, Chapter 3 entailed an awareness to practitioners to recognize possible misunderstandings of applying the ethics of care to business ethics.

Constantly, to frame the implications for practice, the contribution came in the proposition of questions that should be asked for those who design, develop, and deploy AI for its use in firms. Also, throughout the compendium, the contribution was illustrated with examples of the practice (like Amazon firing algorithm, Microsoft chatbot Tay, or algorithms of *learning analytics* and hiring) that explained the specific AI ethical issue in

question and its impact within firms, as well as the way to apply the ethics of care categories and its propositions to ameliorate the issue. All this has implications for how practitioners comprehend their responsibility when designing, developing, and deploying AI.

2. IMPLICATIONS FOR THEORY AND FUTURE RESEARCH

This dissertation contributed to the rising field within business ethics focused on the moral examination of AI. References to the ethics of care approach, have implications on how business ethics scholarship addresses AI ethical impact, and on how the field conceptualizes moral responsibility in this regard.

Also, this dissertation has implications for the design of the curriculum of business ethics education on how it should include the issues explained (such as the problem of moral distance, reinforcing systems of power, unfairly rating and profiling individuals with data, and marginalizing vulnerable stakeholders) and the proposition of the ethics of care to ameliorate those. The same can be said for the curriculum of AI for engineering fields, and in the interdisciplinary studies of AI ethics as an independent discipline.

In general, this thesis entails an extension in the literature of its three main disciplines: business ethics, AI ethics, and the ethics of care. This interdisciplinary approach to the problematics here presented, is unique and needed for the type of problematic and phenomena that this dissertation approaches. However, this interdisciplinary approach also has imitations.

For example, in the propositions of the ethics of care to ethical issues within AI and business, my explanation of the theory and most of my references and propositions referred to what the theory has to offer to AI and Business Ethics. Hence, even though I would have liked to deep more into my explanations of the ethics of care, and present

more authors, and positions, the dissertation focus is on the notions of care that better help AI and business ethical issues. For example, I am not tracing the modern feminist origins of the theory, and I do not devote one chapter to an explanation only of the ethics of care. However, I do explain in chapter three a delineation of the theory that is independent of the two other disciplines of this dissertation, because it remarks needed arguments to avoid possible misunderstandings of the dissertation contribution. Also, I do not devote a chapter just related to AI ethics, or to business ethics, and my references and explanations of the contribution of other moral approaches to AI, in chapter 1, do not deep in the contribution of the approaches apart from AI ethics but are limited to what each approach offer to AI and business ethics.

The said limitations imply that there are still several research paths to undertake. I point out some propositions for future work below. Some of this work may be develop during a possible postdoc at the Technology Ethics Center at the University of Notre Dame. This postdoc will be an extension of my time as a visiting doctoral student.

- <u>What AI should not *touch*</u>

  After all these years of research, I am aware of how in some cases, it might even be irresponsible not to use the power of AI to inform certain decision-making: as in some disease's diagnoses in the medical field. In these cases, if there is a tool to improve the precision of the decision, it is nonsense not to use it. However, I am especially concerned with other types of circumstances when the AI outcome can create an impact, such as the use of the model may exacerbate the harm: as in the case of the COMPAS (used in court to grant or not parole) algorithm where the use of the tool disregards unfair discrimination and even create a new problem

of injustice. The same can be said for most of the examples used in this compendium.

Future research should delimit the scope of AI while investigating if, in some circumstances, we should refrain from the use of this technology. Building on the research of this dissertation, I can propose two arguments for this position, although these are not exhaustive. First, it seems that once an algorithm proposes a path, those who use AI are directly affected by the proposition. For example, there has been the case when an algorithm of face recognition identified someone as guilty, and the police did not contradict the machine and arrest the person (even when the instruction was to use the outcome only as a possibility) (Hill, 2020). This first argument is directly related to the problem identified in Chapter 2 of moral distance and bureaucracy; there the problem of *Hierarchy* which appears in human deference to AI decision-making. Secondly, I argue that, in some context, the impact of the decision or action is so big that there should always be a human in the loop. For this second case take the already explained example of how Amazon is algorithmically rating and firing its drivers without human intervention. Here the drivers only receive an email sent by a bot, telling them they are fired. In this type of context, the impact of the action should be a cause to refrain from the use of AI. Any employee (and person) should be treated with dignity, and the impact of a job loss is so disrupting that it should be done in a certain way: giving voice to the persons affected and treating them in a respectful and attentive way.

These two arguments are not exhaustive, but a proposition to start a conversation of when and where to put limits to the use of AI. Future work should

identify those scenarios where society should refrain from the use of algorithms. Again, without overlooking all the benefits of the good use of AI.

- *Empathy*

  The general problem of moral distance, conceptualized and identified in Chapter 2, explains how AI programs can create a moral distance between those designing and developing AI and the ethical implications of their actions. This dissertation addressed the issue of moral distance proposing to mitigate it from a general approach to the ethics of care. However, future research should focus on the broader understanding of empathy and how empathy can serve as a tool for closing the moral distance in the design and development of AI. The said broader understanding implies connecting different approaches to empathy, like the one of virtue ethics together with the ethics of care, and other focuses related, to undertake the moral abstraction and problems of blurred responsibility of the people related to AI processes.

  For example, scholars in the field of virtue ethics have succeeded in conceptualizing empathy as a cultivated disposition, rejecting the idea of it as an uncontrollable feeling. From this point of view, empathy is a "cultivated openness to being morally moved to caring action by the emotions of other members of our technosocial world" (Vallor, 2016). There, empathy appears as a compassionate concern and "the ability to relate to others while understanding a situation from multiple perspectives" (Ranchordas, 2022). Future research can work to find ways to apply a broader understanding of empathy, considering different approaches (including the ethics of care) to help to ameliorate the problem of moral distance and AI.

Also, the problem of empathy and AI could be addressed related to proximal knowledge and care, addressing the fact that humans are embodied beings and deep in the possible issues of eliminating proximity and embodied knowledge (Hamington, 2004).

Finally, it would be enlightened to deep into the relationship between empathy, partiality, and the ethics of care. This would entail deep into de propositions of chapter 3, which stated that the ethics of care is not about partiality, for example, the kind of partiality based on feelings of pity.

- *The problem of many hands and AI*

  The problem of many hands, "described as the problem of attributing or allocating individual responsibility in collective settings," (van de Poel and Zwart 2015) has been explained in Chapter 2 as a problem of moral distance and then implied to AI. However, specialized work is still needed to conceptualize this problem and its moral implications. Empirical work here would be illustrative.

- *Trade-offs and good practices*

  Lastly, while finalizing this dissertation, I particularly worry about how firms keep framing AI ethical issues in terms of trade-offs (Telkamp and Anderson, 2022). Take the example of data privacy and how the general discourse is how data work like currency to obtain technological services (in social media, web services, email services, and the like), sometimes giving enterprises the "right" to violate rights of privacy. In that case, the argument is that if personal data is not the currency, all those services would be charged, which may create problems of inequality. This argument was constantly referred to in the Seminar *Digital*

Footprint of Fundación Pablo VI that I attended as a report writer. Representatives of private enterprises, such as insurance companies, continually referred to trade-offs and how society should sometimes accept things to favor efficiency (see Fundación Pablo VI[56]). Future research should address this problem while proposing best practices to favor efficiency but also limiting the scope of AI. For example, there may be the case that it is not necessary to speak of an "all or nothing," and companies can use our data in some cases, well defined, but in others may be prohibited. Here (for example) it would be possible, perhaps, to use our data for particular advertisements, but not to exploit vulnerabilities and deliberately harm individuals, denying them opportunities to develop and obtain services.

REFERENCES

Hamington, M. (2004). Embodied care: Jane Addams, Maurice Merleau-Ponty, and Feminist Ethics. University of Illinois Press.

Hill, K. (2020). Wrongfully acussed by an algorithm, *The New York Times*. Retrieved from: https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html

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

Ranchordas, S. (2022). Empathy in the digital administrative state, 72 *Duke Law Journal* (forthcoming).

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*.

[56] https://www.fpablovi.org/sintesis-huella-digital

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

van de Poel, I and Zwart ,S.D. (2015). Conclusions. From understanding to avoiding the problem of many hands. In van de Poel, I., Royakers, L., Zwart, S.D. (Eds) Moral Responsibility and the Problem of Many Hands. Routledge, New York.