

## **ETHICS OF CARE AS MORAL GROUNDING FOR AI**

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 Commission, 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 of 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.

## **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. 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.

## ETHICS OF CARE

Ethics of care appeared as a theory in the XX Century. 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 inattentive, responsive, and respectful ways.” This definition encompasses what implies going beyond the formal rationality based on principles, guidelines, and norms within ethics.

## 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.”<sup>1</sup> Those that are marginalized or do not fit a particular pattern are then denied rights or further harmed feel that algorithmic

---

<sup>1</sup> Retrieved from: <https://mimionuoha.com/algorithmic-violence> (July, 2021).

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.

## 1. 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?

## 2. Relationships and interdependence

In 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:

- 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?

### 3. **Direct Level**

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?

### 4. **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 prevent 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?

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

Floridi, L., Cows, 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., Cows, 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>