Coping with Artificial Intelligence Ethical Dilemma and Ethical Position Choices?

Sylvie Gerbaix[1] [a], Sylvie Michel[2] [b] and Marc Bidan[3] [c]

¹Laboratoire MRM, Montpellier University, Aix-Marseille University, France
²Univ. Bordeaux, IRGO, UR 4190, F-33000 Bordeaux, France
³LEMNA, Nantes Université, France

Keywords: Artificial Intelligence (AI), Algorithmic System, Ethical Theory, Bias, Responsibility, Transparency.

Abstract: The aim of this conceptual article is to demonstrate that proposing measures, actions, and decisions to improve the ethics of Artificial Intelligence (AI) depends on the ethical theoretical position chosen. To achieve this, we proceeded in two stages. Firstly, we characterized and synthesized three different ethical issues posed by AI. Secondly, we selected two main ethical positions proposed by philosophical literature. Finally, we showed that the choice of an ethical theoretical position for each category of ethical issues of AI leads to different decisions. We demonstrated that for each category of ethical problems, the ethical decisions and their consequences differ depending on the ethical theory chosen. The value of this paper is to highlight that the literature on AI ethics often neglects the implications of choosing an ethical position. In order to attempt to solve ethical issues, it is necessary to reach agreements and have discussions that take into account the different ethical theoretical positions and their consequences in terms of decision-making.

1 INTRODUCTION

Today, AI helps us in selecting footage, music, friends, and partners (Milano et al., 2020). It also supports institutions in making legal decisions, maintaining public order, and helps doctors in providing a diagnosis (Obermeyer, 2019), traders in trading (Aggarwal, 2021), and armies in using killer robots to achieve their goals. Numerous areas seem to be under the yoke of AI capacities, including logistics, health, education, research, defense, banking, agri-food, culture, leisure, social, and professional networks. Since algorithms are at the heart of human relations and exchanges, questions relating to uses (and misuses), benefits (and limits) have become crucial. Faced with this surge of artificial intelligence, the ethical question is urgent.

Artificial intelligence covers a wide range of research and computer applications, including machine learning, computer vision, knowledge representation, language processing, and decision support. With the notion of AI, we propose that it is not the algorithm alone that can be problematic, but rather its embeddedness in a system, a set of actors, power norms, and complexity (Neyland, 2016; Seaver, 2017). Thus, in this contribution, we consider AI not as purely technical objects, but as technical systems embedded in culture(s) and which can be seen, used, and approached from different perspectives (legal, technological, cultural, social). It is a technical construction that is both deeply social and cultural. It does not escape social and cultural construction like all other tools, neither in its development, nor in the inputs (data), nor in the interpretation of the results (output), nor in its use.

Some authors (Hamet & Michel, 2018) have shown that the ethical questions that arise in information systems are specific to this field, in the sense that the ethical dilemmas posed by AIs do not arise, or do not arise in the same way in other fields. However, even if some questions appear to be specific to AI, it seems necessary to have a theoretical framework to study them and provide answers.

[a] https://orcid.org/0009-0005-3544-2399
[b] https://orcid.org/0000-0002-8175-9996
[c] https://orcid.org/0000-0003-1739-5697

Numerous articles today attempt to analyze the main ethical problems linked to AI by proposing, for the most part, courses of action, standards, and codes to be put in place (Mittelstadt et al., 2016; Berreby et al., 2017; Anderson et Anderson, 2018; Yu et al., 2018; Buhmann et al., 2020; Tsamados et al., 2021). However, we note that these studies are not part of a theoretical ethical current. We would like to show that when faced with an ethical problem raised by AI, the recommendations and decisions may not be the same depending on the initial theoretical ethical position. To this end, we will first present two main theoretical ethical currents that we consider important: Kant's deontological ethics and Hegel's consequentialist ethics. A synthesis of the ethical questions raised by AI is then presented. In a third point, we confront them with the two theoretical ethical positions. We then show that the decisions to be taken may differ. Thus, we show that the norms to be put in place in the face of the ethical problems of AI will only be effective if a theoretical ethical position is chosen ex ante.

2 SYNTHESIS OF ARTIFICIAL INTELLIGENCE (AI) ETHICAL ISSUES IN THREE THEMES

2.1 AI and Responsibility

Here we need to point to De George (2009) who highlights an unfortunate trend of giving up the assignment of responsibilities when it comes to Information Systems or algorithms, a renunciation that comes from two sources. The first is what De George (2009) refers to as the myth of amoral Information and communication technology (ICT). This myth amounts to limiting ICT to their technical aspect and to considering that machines cannot, obviously, be held responsible for the consequences of their use. There is an avoidance of responsibility through "the computer said so" type of denial (Karppi, 2018). The second source of dilution of responsibility comes from a split, in decision-making, between the developers, who believe they are fulfilling their duty by strictly respecting the requests they receive, and the management, who does not consider itself responsible for the technology-related flaws. However, these developers are never face to face with the stakeholders which creates a new ethical problem linked to the distance in decision-making (Rubel et al., 2019). This issue of responsibility is very current, particularly with the

development of AI (Martin, 2019; Yu et al., 2019, Wieringa, 2020).

Faced with this dilution of responsibilities, some authors (Chatterjee et al., 2009; Light and McGrath, 2010) call for the adoption of a disclosing ethic approach considering the algorithms as an actor in its own right, and aiming to reveal the ethical questions posed by their design, and not just their use. Machines should be able to make ethical decisions using ethical frameworks (Anderson et Anderson, 2018). Davison (2000), Stahl, (2004) or Reddy (2019) note the importance of assigning responsibility, especially given the seriousness of the potential consequences of an error. Some authors suggest identifying levels of responsibility (individual, hierarchical, collective and organizational) or suggest assigning responsibilities according to the role of stakeholders, according to their decision-making (Chander, 2017; Kraemer et al., 2011; Torresen, 2018). Others (Chander, 2017; Kemper and Kolkman, 2019; Buhmann et al., 2020) stresses that an organization should take responsibility for their algorithms regardless of how opaque they are (Malhotra et al., 2018).

2.2 AI and Bias

The algorithms that work on language are fed by billions of data (texts, images, videos...) steeped in our cultures. When a system becomes expert enough to simulate a conversation and produce language that sounds natural, it relies on the commonly accepted ideas of the society it is addressing. Not surprisingly, it reproduces ethically questionable historical cultural representations. Also, for example, from the 2000s, a new and important question emerged in the literature, under the term of social sorting (Hamet and Michel, 2018). This trend, stemming from surveillance studies, examines the risk that the analysis of personal data will lead to segregation or discrimination. Employers however no longer hesitate to use tri- social algorithms to recruit. Banks and insurance companies conduct scoring policies based on these social sorting technologies. Real estate agencies and social landlords also carry out de facto discrimination by establishing choices of allocation or not, using housing algorithms which are based on the last name, on first name, on address, on the mastery of the French language, etc. If people tend to recruit fewer women, the algorithm will implicitly reproduce this trend. This is a crucial issue with AI and deep learning algorithms with which a machine is able to learn through its own data processing. Thus, biases in AI, based on masses

of previous data, are likely to result in discrimination and exclusion, on a scaled-down scale by reproducing prejudices incorporated in unprocessed data. AI, combined with the processing of massive data, even induces an autonomous functioning of processing social characteristics. It is the self-learning AI itself that produces and reproduces "social sorting". In this sense, AI did not invent discrimination, but they participate in this movement, by reproducing it, even by intensifying it.

2.3 AI and Transparency

Due to the multiplicity of actors, stakeholders and links, algorithms, are increasingly opaque. However, it is necessary to open the black box of algorithms, to offer better transparency (Guidotti et al., 2020; Buhmann et al., 2020) and then possible traceability. As Turilli and Floridi (2009) note, transparency is not an "ethical principle in itself but a pro-ethical condition for enabling or impairing other ethical practices or principles" (p.105). This is why it is important to distinguish between the different factors that may hinder transparency of algorithms, identify their cause (Diakopoulos and Koliska, 2017). This transparency is also a request from individuals in relation to the protection of their privacy, the confidentiality of stored data and surveillance. This search for transparency involves being able to clarify and decipher multiple involved and intertwined processes, processes such as the collection of data and their validation, the action and decision-making processes of the chain of actors in the realization of algorithms (initial decision, project funders, AI researchers, analysts, developers), decision-makers and funders of uses (for example, for health systems or autonomous cars) (Martin, 2019). Also, the ethical problem of the transparency of algorithms can lead us to broader ethical questions. For example, as early as the 1960s, some worried that advances in science and technology could threaten the functioning of democracy, because only a few experts were able to truly understand complex technologies (Habermas, 1970). The technocracy hypothesis arose through portraying a future society where experts would make decisions based on their own value system, offering their best solution. Then, after the experts, the AI could also lead to a form of algorithmic governmentality based on the statistical processing of data communicated (voluntarily or without their knowledge) by citizens, in particular thanks to sensors connected, leading to a pre-elaboration of the collective decisions thus elaborated by the algorithms.

3 ABOUT TWO MAJORS ETHICAL CURRENTS

MIT has initiated the Moral Machine Project, which leverages the collective wisdom of crowds to devise solutions to ethical dilemmas related to autonomous vehicles controlled by AI that could potentially cause harm to pedestrians and/or passengers in case of malfunction. While the wisdom of the crowd is used as an ethical reference in this project, one could alternatively appeal to the categorical imperative, the concept of virtue, the consequences of actions, or other norms. Therefore, we believe it is essential to present the major ethical theories at hand to properly address ethical issues (Hamet and Michel, 2018). Hence, we will discuss two primary ethical movements, namely Kant's deontological ethics and Hegel's consequentialist ethics. While we acknowledge that this is not an exhaustive list, these two ethical currents can demonstrate the significant challenges encountered in addressing the ethical problems posed by algorithms.

3.1 Kant's Deontological Ethics

During the 18th century, Kant, in his work Critique of Practical Reason (1788), attempted to answer the question, "What shall I do?" This is Kant's second question (after "what can I know," which is dealt with in the Critique of Pure Reason), and the third is "what can I hope for," which is dealt with in the Critique of the Faculty of Judgement. At the beginning of the Critique of Practical Reason, Kant asks whether it is possible to construct a moral rationalism, a supreme principle of rationality which would be the moral law. He transforms the question "what shall I do" into "what are the supreme principles of morality." In this work, he promotes the autonomy of the will. It is no longer up to the human will to align itself with the good as an external standard, but rather it is up to the will to define the good as that which is universally desirable. Kant sets out to find "practical laws," i.e., "objective principles valid for the will of every reasonable being." To be objective and universal, moral obligation must be expressed by a formal principle, an a priori, universal, and necessary criterion. It is therefore the autonomy of the will that must constitute the sole principle of all moral laws and duties. The fundamental law of pure practical reason can be stated as follows: "Act only according to the maxim that you can will at the same time that it becomes a universal law."

3.2 Hegel's Ethics

Unlike Kant, Hegel relies on the experience of human beings and aims to understand what makes morality, rather than looking for where it should come from.

3.2.1 Hegel's Critique of the Kant's Moral Law

In The Principles of Legal Philosophy, in its second part, Hegel addresses several criticisms of Kant's ethical principles and his moral law. In short:

First, the moral law is dolorist: it always presupposes suffering in its execution, in the sense that one must do with aversion what duty dictates. According to Hegel, one can do one's duty with pleasure.

Second, circumstantial context is not taken into account. This means that moral duties are conditioned by the situation. Duties are not immediate; they require reflection. Most norms certainly work as a habit, which frees the mind, allowing one to think about when it is necessary to take the situation into account. There is a hierarchy of norms. In the famous example of lying, for Hegel, lying is seen as the least bad solution in a given context.

Third, the moral law presents a theory that is ahistorical. This is the most problematic point for Hegel: Kant does not take into account the evolution of norms and society. Hegel would say that morality has something provisional, that it evolves over time (referring to Descartes and his provisional morality).

So, Hegel proposes a theory of action and moral imputation: what makes an individual's morality is their actions. What makes the morality of an action is norms and morality. What makes it possible to impute the action to the individual is ethics.

3.2.2 Hegel's Concept of Ethical Life

In the third and final part of The Principles of the Philosophy of Law, Hegel develops his concept of ethical life, which encompasses the family, civil society, the state, and world history (§142 to §360). Ethical life is the set of norms that emanate from the institutions that regulate the life of people: the family, the corporations, and the state, which is the basic institution and the condition of possibility for other institutions. In ethical life, there are obligations, which today are called norms.

4 AI ETHICAL ISSUES AND ETHICAL POSITIONS CHOICES

4.1 Lack of Ethical Positioning in the Literature

Recent literature on these ethical problems has proposed some interesting avenues for reflection and action. For example, on the issue of fairness, some authors recommend to develop a sociotechnical framework to address and improve the fairness of algorithms (Edwards and Veale, 2017; Selbst et al., 2019; Wong, 2019; Abebe et al., 2020). Concerning the issue of responsibility, for example, Shah (2017) analysis points out that the risk of some stakeholders failing to meet their responsibilities can be addressed, for example, by creating separate bodies for the ethical oversight of algorithms. However, others show that expecting a single oversight body, such as a research ethics committee or institutional review board, to be 'solely responsible for ensuring ethical rigor, utility and probity is unrealistic (Lipworth et al., 2017). Concerning the issue of transparency, for example, Gebru et al. (2020) propose that the transparency constraints posed by AI can be resolved, in part, by using standard documentation procedures similar to those deployed in the electronics industry. In addition, another recent approach is the use of technical tools to test and audit AI and decision making. This involves checking algorithms for negative trends, such as unfair discrimination, and auditing a prediction or decision track in detail, (Weller, 2019; Malhotra et al., 2018; Brundage et al. 2020). We do not question these courses of action, but we insist that in order to choose one of the proposed courses of action, it is first necessary to be part of an ethical current, as the answers differ, as we will show in the following.

4.2 Different Answers to Ethical Questions Depending on the Positions Chosen

AI are loaded with values (Brey et Søraker, 2009; Kraemer et al., 2011; Mittelstadt et al., 2016; Tsamados et al, 2021) contradicting the myth of the long-lived neutral algorithm. Indeed, any algorithm involves a multitude of decisions, whether classification, prioritization, display, filtering, learning. However, the choice of filtering techniques, classified data, the options chosen may well reflect a certain understanding of the world. AI

are nothing more than ideas, opinions formalized in code, and in no way escape the subjectivity of developers, managers, contractors or society. Algorithmic coding contains a wide spectrum of standards that can range from moral injunctions to unconsciously integrated norms. We want to show that choosing these solutions without first choosing an ethical position is a bad approach. Indeed, the solution chosen can only depend on the ethical position initially taken. Responses to these ethical problems will differ depending on whether one adheres to a deontological Kantian ethics or a consequentialist Hegelian ethics.

With regards to the ethical theme of biases, according to Kantian's ethics, AI in its entirety must respond to a categorical imperative of non-discrimination and justice. AI must not discriminate or reproduce discriminations. This could be one of the first categorical imperatives assigned to this ethical theme. It would then be a question of knowing how to put in place this ethic approach to duty. Consequentialist's, on the other hand, will be interested in the consequences of these biases and discriminations with regards to the well-being of society. Some discrimination can be accepted. They may consider that discrimination against a minority is not detrimental to social well-being and therefore that it is acceptable.

With regard to the theme of responsibility, we find the two positions defined above. For Kantians Ethics, faced with the tram dilemma type (thought experiment which offers a person a choice of action, knowing that if he acts, his gesture will benefit a group of people, but will harm a person), dilemma which is very similar to the dilemmas posed to the autonomous car, it is not possible that a person will be killed because of my action. For consequentialists ethics, faced with this dilemma, between one or more people, between young or old, depending on the social utility of the person, we can make a choice. Choosing to sacrifice one person to save five can be understood according to this morality.

Finally, regarding the theme of transparency, the ethical imperative leads us to consider the initiator of the standard - whether it's the machine or the human. Is it ethical to delegate decision-making to a machine? Can a machine have good intentions? According to Hegelian ethics, we must identify the consequences of a loss of control and human mastery. The time, financial, and fatigue savings generated by delegating decision-making to a machine can justify certain costs, such as the loss of certain degrees of freedom and generalized surveillance. On the other hand, Kantian ethics

emphasizes individual reflection. In order to create contextualized ethical norms, Hegel proposes three levels of action: family, civil society, and state. We can summarize our approach in the following table1:

Table 1.

AI Ethical questions - Synthetic formulation	Questions raised according to ethical theorical positions
Bias Questioning the integration of societal values in AI. The rules included in the algorithms are not neutral and convey conscious and unconscious values of the developers, the organizations, the companies.	**Kantian ethics:** questioning of Universal moral rules to be integrated into algorithms (absence of discrimination) and to be imposed on society. Fairness couldn't' be accepted, never. **Hegel ethics** and consequentialists current: questioning the consequences of the biases generated. Some on theses could be accepted.
Responsibility Amoral AI; Identification of responsibilities? Dilution of responsibilities; Revealing ethics and the system as an actor in its own right	**Kantian ethics:** questioning a "categorical imperative" (the autonomous car must not kill, for example). Responsibility must always be accurately attributed **Hegel ethics** and consequentialists current: questioning the consequences of ethical dilemmas. A trade-off between different responsibilities is possible
Transparency Transparency of the process: data/algorithmic processing/ effects Traceability	**Kantian ethics:** questioning the initiator of the categorical imperative (Human or algorithm) **Hegel ethics** and Consequentialists current: questioning the consequences of the excesses of the loss of human control

5 CONCLUSIONS, DISCUSSIONS AND PERSPECTIVES

In summary, in the proliferation of ethical problems posed by AI, three paths of ethical questioning related to AI have been identified: the question of algorithmic biases, the question of responsibility,

the question of transparency. We show that these themes need to be according to the ethical theory in which they are questioned, (Deontological, Consequentialists).

The contributions of this article are therefore twofold. First, we have consolidated the literature on the topic of the ethics of AI, by putting forward a synthesis of the main ethical questions. This synthesis sheds light on this confused field. We have also shown, with support of the main ethical theoretical currents, that these questions can lead to very different answers. Thus, in terms of practical ethics, it is not enough for organizations to identify ethical issues and propose ethical charters. The first task, which underlies the whole, is indeed to choose an ethical theoretical position.

REFERENCES

Abebe R., Barocas S., Kleinberg J., Levy K., Raghavan M., Robinson D.G., (2020) Roles for computing in social change, *in Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 252–260.

Aggarwal N., (2021), The norms of algorithmic credit scoring, *The Cambridge Law Journal, vol. 80, n°1, pp. 42–73*.

Anderson, M., & Anderson, S. L. (2018). GenEth: A general ethical dilemma analyzer. *Paladyn, Journal of Behavioral Robotics, 9(1), 337-357.*

Berreby F., Bourgne G., Ganasc J.G., (2017) A declarative modular framework for representing and applying ethical principles, *In 16th Conference on Autonomous Agents and Multiagent Systems*, 2017, May.

Brey P., Søraker J.H.,(2009) Philosophy of computing and information technology. *In Philosophy of technology and engineering sciences, pp. 1341–1407, North-Holland.*

Brundage M., Avin S., Wang J., Belfield H., Krueger G., Hadfield G., Khlaaf H. et al., (2020) Toward trustworthy AI development: mechanisms for supporting verifiable claims, *arXiv preprint arXiv:2004.07213.*

Buhmann, A., Paßmann J., Fieseler J.C., (2020) Managing algorithmic accountability: Balancing reputational concerns, engagement strategies, and the potential of rational discourse, Journal of Business Ethics, vol. 163, n°2, pp. 265–280.

Chander A., (2017) The racist algorithm?, Michigan Law Review, vol. 115, n° 6, pp. 1023–1045,.

Chatterjee S., Sarker S., Fuller M., (2009) Ethical information systems development: A Baumanian postmodernist perspective, *Journal of the Association for Information Systems, vol. 10, n°11, pp. 787–815.*

Davison R.M., (2000) Professional ethics in information systems: A personal perspective, *Communications of the association for Information Systems, vol.3,n°1.*

De George R.T. (1999), Business ethics and the information age, *Business and Society Review, vol 104, n°3, pp. 261–278.*

Diakopoulos N., Koliska M., (2017), Algorithmic transparency in the news media, *Digit Journal, vol 5, n°7, pp.809–828.*

Edwards L., Veale M., (2017) Slave to the algorithm? Why a right to explanation is probably not the remedy you are looking for, *SSRN Electron J.,*

Gebru T., Morgenstern J., Vecchione B., Vaughan J.W., Wallach H., Lii H.D., Crawford H,K., (2020) *Datasheets for datasets, Communications of the ACM, vol. 64, n° 12, pp. 86–92, pp.*

Guidotti R., Monreale A., Ruggieri S., Turini F., Giannotti F., Pedreschi D., (2018) A survey of methods for explaining black box models, *ACM computing surveys (CSUR), vol 51, n°5, pp. 1– 42, 2018.*

Habermas J. (1970), Towards a theory of communicative competence, *Inquiry, vol. 13, n°1–4, pp. 360–375.*

Hamet J., Michel S., (2018), Les questionnements éthiques en systèmes d'information, *Revue française de gestion, vol. 44, n°271, pp. 99–129.*

Kant E., (1905 ,orig 1787) *Critique de la raison pratique*, Editions Flammarion.

Karppi T., (2018), The computer said so: on the ethics, effectiveness, and cultural techniques of predictive policing, *Soc MedAI Soc, vol 4, n° 2,.*

Kemper J., Kolkman D.,(2019), Transparent to whom? No algorithmic accountability without a critical audience, *Information, Communication & Society, vol 22, n°14, pp. 2081–2096.*

Kraemer F., Van Overveld K., Peterson M., (2011), Is there an ethics of algorithms?, *Ethics and information technology, vol.13, n°3, pp.251–260.*

Light B., McGrath K., (2010), Ethics and social networking sites: a disclosive analysis of Facebook, *Information Technology & People.*

Lipworth W., Mason P.H., Kerridge I., Ioannidis J.P.A., (2017), Ethics and epistemology in big data research, *J Bioethical Inq, vol. 14, n°4, pp.489– 500.*

Malhotra C., Kotwal V., Dalal S., (2018), Ethical framework for machine learning, *ITU Kaleidoscope: machine learning for a 5G Future, pp. 1–8. Santa Fe: IEEE,.*

Martin K., (2019), Designing ethical algorithms, *MIS Quarterly Executive June, 2019.*

Milano S., Taddeo M., Floridi L., Recommender systems and their ethical challenges, *Ai & Society, vol.35, n°4, pp.957–967, 2020.*

Mittelstadt B.D., Allo P., Taddeo M., Wachter S., Floridi L., The ethics of algorithms: Mapping the debate. *Big. Data & Society, vol.3, n°2; 2016.*

Neyland D., (2016*),* Bearing accountable witness to the ethical algorithmic system, *Science, Technology & Human Values, vol.41, n°1, pp.50–76.*

Obermeyer Z., Powers B., Vogeli C., Mullainathan S., (2019), Dissecting racial bias in an algorithm used to manage the health of populations, *Science,vol.366, n°6464, pp. 447–453,.*

Reddy E., Cakici B., Ballestero A., (2019), Beyond mystery: putting algorithmic accountability in context, *Big Data Soc, vol.6, n°1.*

Rubel A., Castro C., Pham A., (2019), Agency laundering and information technologies, *Ethical Theory Moral Pract, vol. 22, n°4, pp.1017–1041.*

Selbst A.D., Boyd D., Friedler S.A., Venkatasubramanan S., Vertesi J., (2019), Fairness and abstraction in sociotechnical systems, *Proceedings of the Conference on Fairness, Accountability, and Transparency, pp. 59–68. Atlanta, GA, USA: ACM Press,*

Shah H., (2018), Algorithmic accountability, *Philos Trans R Soc A: Math Phys Eng Sci, Vol 376.*

Seaver N., (2017), Algorithms as culture: Some tactics for the ethnography of AI, *Big data & society, vol. 4, n°2*

Stahl B.C., (2004), Information, ethics, and computers: The problem of autonomous moral agents, *Minds Mach, vol. 14, n°1, pp. 67–83, 2004.*

Torresen J., (2018), A review of future and ethical perspectives of robotics and AI, *Frontiers in Robotics and AI, vol. 4, pp. 75.*

Tsamados, A. , Aggarwal, N. , Cowls, J. and Morley J., and Roberts H., Taddeo M., Floridi, L., (2021), The Ethics of Algorithms: Key Problems and Solutions *37 AI & Society 215*

Floridi L., (2021), The ethics of algorithms: key problems and solutions, *AI & SOCIETY, pp. 1–16.*

Turilli M., Floridi L., (2009), The ethics of information transparency, *Ethics and Information Technology, vol.11, n°2, pp.105–112*

Weller, A. (2019), Transparency: motivations and challenges. In *Explainable AI: interpreting, explaining and visualizing deep learning* (pp. 23-40). Cham: Springer International Publishing.

Wieringa, M., (2020), What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability, *In Proceedings of the 2020 conference on fairness, accountability, and transparency, pp. 1-18, January 2020.*

Wong P.H., (2019), Democratizing algorithmic fairness. Philos Technol,.

Yu, H., Shen, Z., Miao, C., Leung, C., Lesser, V. R., & Yang, Q. (2018). Building ethics into artificial intelligence. arXiv preprint arXiv:1812.02953.