



HAL
open science

Artificial Intelligence: A Tale of Social Responsibility

Cécilia Darnault, Titouan Parcollet, Mohamed Morchid

► **To cite this version:**

Cécilia Darnault, Titouan Parcollet, Mohamed Morchid. Artificial Intelligence: A Tale of Social Responsibility. 2019. hal-02270410v1

HAL Id: hal-02270410

<https://hal.science/hal-02270410v1>

Preprint submitted on 25 Aug 2019 (v1), last revised 25 Oct 2019 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Artificial Intelligence: A Tale of Social Responsibility

Cécilia Darnault¹, Titouan Parcollet^{2,3}, Mohamed Morchid²

¹Avignon, France

²LIA, Avignon Université, France

³Orkis, Aix-en-provence, France

cecilia.darnault@gmail.com

Abstract

Conversely to the legislation that struggles to develop, regulate and supervise the use of artificial intelligence (AI), the civil society, that gradually realizes the fundamental issues and perspectives induced by this new technology, slowly starts to take responsibility and to mobilize. Social responsibility expresses itself through the emergence of new voluntary standards, that could integrate the concept of *social good* with the use of AI. More precisely, this paper proposes to develop three axes of tools for the social responsibility in AI, including stakeholder awareness, the integration of ethical and technical standards to induce good behaviors, and the incitement to a responsible AI.

1 Introduction

Novel technologies and especially Artificial Intelligence (AI) paradigms induce a massive interest from researchers from different domains and raise questions and concerns. Indeed, some “*worry about the transformations and the possible destruction that could put in danger our world*” [19, 25], while “*others, convinced of the inevitability of the upheavals to come, seek to influence the movement to make this future livable*” [19]. The scientific community does not remain indifferent and warns AI based systems users against the risks related to AI models and the manner that the models have been learned. These concerns are shared by the society [34, 18], but are also difficult to develop and formalize. In spite of few recent specialized applicable standards, especially concerning the protection of personal data and privacy, or in robotics, a general and global legal system to control the risks related to AI, does not exist. Fortunately nature abhors a vacuum and solutions have been developed. In fact, at the edge of lawlessness and mandatory law, a whole range of applicable standards more or less restrictive are emerging to fill possible legal loopholes, often out of necessity, as prior regulatory basis or to overcome political and economical issues at a national or international scale [21, 28]. In the same manner as the development of private initiatives, and with non-binding legal tools also known as *softlaw* [16, 11] that are actively used in other legal domains, such as the recent environmental law, the civil society is mobilizing to guarantee the development of AI for social good. The term *responsibility* defines the principle of facing the consequences of our actions. However, the responsibility is expressed in different ways such as a coercive normative constraint, an economic mechanism, a moral imperative or a governance mechanism [13, 12]. Consequently, a social accountability emerges and leads to the deployment of voluntary standards coming from the civil society to guarantee a suitable principle of *social good* in addition or as a precursor to the traditional normative framework.

This paper explores and defines **how voluntary standards enable the use of AI for social good** with the introduction of different tools (Figure 1 in Appendix) that have been successfully employed in other legal areas such as the environmental law. The major forms of voluntary standards applicable to AI to ensure its positive use according to the social good are gradually introduced, starting with the stakeholder awareness (Section 2), followed with the voluntary application of private standards (Section 3), and concluding with the incitement to the development of AI for social good (Section 4).

2 Awareness on social good for AI

The actors empowerment and the integration of social good for AI start with the democratization of the information related to both AI risks for social good, and to the awareness of the concepts of social good for AI.

2.1 Risk awareness of AI for social good

The fast development of AI leads to concrete risks to the social good in different aspects of our societies. Different examples are available such as in the economic field, with the strong impact of AI on robotics and the employment market, or in the social area with the unfortunate transfer of important human biases to the systems (*e.g.* racism, sexism, etc.) [37, 36, 9], or within the security domain with concerns involving the relations of AI to the concept of democracy, arms monitoring or in a legal aspect, with multiple questions on the protection of personal data and privacy, or on the problem of accountability of AI systems [32, 33]. Raising the awareness of AI risks is of crucial interest to ensure its implementation with respect to the standards of social good. Without a normative framework, and in the same manner as the recent mobilization of numerous employees against a partnership involving their AI products and the army [5], some researchers become aware of these risks, and take responsibility and fight to establish safeguards for a responsible development of AI [7, 5]. A growing awareness expressed in the economic sphere by the recent creation of ethical committees from private companies, such as the *advanced technology external advisory council* or the foundation on an ethical institute, aiming both to conduce independent researches increasing awareness by providing advises on the subject of AI for social good [4, 2]. Moreover, and at the country level, some governments realize the lack of standards and the increasing risk of AI, and start to build national study committees. As an example, France and Canada have recently announced the joint creation of the *international panel on artificial intelligence (G2IA)* [3]. The latter mechanisms are expected to promote a respectful and ethical approach of AI with the principles of sustainable development. Unfortunately, some of them already suffer from internal biases on the subject [4] highlighting the complexity of the task. Concurrently to these private initiatives, numerous open events promoting a risk awareness on AI for social good emerge, including well-known international conferences and workshops broadening the debate of stakeholders on this topic. These mechanisms are crucial to raise the risk awareness on social good for AI and further efforts have to be put to improve their impact.

2.2 Raising the AI actors awareness to the social good

Well-known researchers on social sciences and on the management theory I. Nonaka and H. Takeuchi have proposed an interesting approach particularly suited for these drawbacks called the rugby metaphor. The latter idea highlights the richness of the information that can be cultural, moral, emotional or technical as well as the importance of the selection and transfer processes of the information by the actors of a specific situation [8]. This work demonstrates the importance of the information and of its circulation among the various stakeholders, but most importantly, it shows the need to mobilize relevant and meaningful information at the right moment. Based on these hypothesis, the use of AI for social good necessarily involves both formal and informal education and training to allow a transfer of the relevant information to raise the awareness of the different actors. In the same manner as computer science students are often introduced to the economics and management sciences, AI actors could and have to be initiated to fundamentals of social good (*e.g.* basic rights, social sciences, risk management, etc.) to integrate them whether consciously or unconsciously to their research or processes of production. Conversely, a dissemination of non-scientific or basic knowledge on methods related to AI to the civil society is crucial and must be encouraged to simplify the integration of the AI-related concepts and their impact on our societies. As an example, it is feasible to apply existing tools, including advertising campaigns or public seminars, to leverage a minimal but global knowledge on AI. Then and as a result of corporate social responsibility, some organizations have set up an entire legal culture that could be transferred to AI to develop a collective intelligence of their teams regarding AI. A very simple, but powerful tool, is to create and broadcast informative e-learning videos.

3 Private standardization of social good for AI

Once the different actors of AI are well-aware of the concept of social good, various instruments can be proposed to formalize its integration and induce behaviors that are respectful of the idea of social good within the use of AI, such as the development of voluntary standards from private initiatives.

3.1 Ethical standards for social good in AI

As stated by the Nobel Prize for Economics Joseph Stiglitz, “ *we are a global community, and like all communities, we must follow rules to live together. They must be clearly seen as fair and just. They must pay due attention to the poor as well as the powerful, and demonstrate a deep sense of honesty and social justice*” [17] and regardless of the considered domain. In this extent, AI constitutes a major scientific and technological breakthrough that brings important social benefits, but also rises critical ethical and social risks [7]. Ethical challenges that the community, and in the absence of a legal framework in this area, has took off. Indeed, among the tools of voluntary standardization to enhance the concept of social good within AI domain, some ethical standards have recently emerged while others are still in development. As an example, the *Montréal declaration for a responsible development of artificial intelligence* [7] is advocating for a positive development and use of AI following specific social good principles. Following the latter declaration, others institutions such as the *Council of Europe* are interested in extending these principles to a general ethical framework for a responsible and ethical use of AI with respect to the Human Rights, rule of law, and democracy. More precisely, the *Council of Europe* investigates the impact of AI in specific domains, including the medical field (Bioethics Committee) [35], with a strategic action plan on the interactions of technologies and Human Rights in the biomedical context, or gender equality (Gender Equality Committee), with an funded project to prevent and reduce the risks of sexism induced by AI algorithms [37, 36, 9]. In fact, numerous areas with an emphasize on AI are investigated, including education, discrimination, cyber-security,.... Despite various ongoing projects on ethical standards that constitute a significant progress toward the supervision of AI for social good, coordination and the practical application of these standards remain an open and crucial problem. Indeed, all these institutions could benefit from coordinating their ethical standards. More precisely, a global coordination would enable an international ethical framework for AI, and would make the effectiveness of its application easier. Furthermore, and due to the abstract concepts driven by the idea of an ethical AI, an effective implementation of these texts implies to improve them with more global actions. As an example, we propose to conceive specific ethical “*codes of conduct*” for professionals, to deploy in private companies or in public institutions.

3.2 Technical standards for social good in AI

In the last decades, the concept of social responsibility, that relies on the voluntary initiative, has been highly investigated in the field of corporate social responsibility (CSR) for the environmental protection [22, 24, 14]. The social responsibility answers a global need for references (*e.g.* standards, laws), enabling an institution or a company (*e.g.* AI-related products) to fully integrates its economical, social and legal environments alongside with its different stakeholders, within the management of the institution or company. The standardization is a crucial element of the *softlaw*, that represents common standards to illustrate and standardize the practices of the social responsibility, and ensures a certain efficacy by proposing a collective solution to technical or organizational issues [10, 23]. Various national, regional or international organizations are at the origin of these voluntary standards. In fact, and in the same manner as *W3C* that has standardized the compatibility of certain web-based technologies, a few organizations start to embrace the concept of AI and its supervision for social good. As an example, the *International Standardization Organization* (ISO) is currently working toward new strong standards for AI. Despite the fact that recently published standards mostly consider the information technologies (IT) and the reference data architectures, numerous ongoing or upcoming projects focuses on the use of AI, such as the standard *ISO/IEC AWI 38507* that relates to the governance of IT and the governance impacts of the use of AI technologies [6]. In the same context, a French standardization association, named AFNOR, actively works on important aspects of the use of AI including ethical and social concerns, reliability and the integration of the risks induced by systems relying on machine learning based solutions [1].

4 Incitement to social good for AI

Based on the previous introduction to social behaviors of social good standards for the use of AI, it is now feasible and needed to actively promote their development toward the community and to induce a generalization of social good in AI.

4.1 Contractual commitment to social good for AI

While most of the contractual relations are limited to the principal intentions of the parties and solely consider their main duties [27], it is worth emphasizing that a contract not only establishes a commercial relationship but also enables to anticipate foreseeable risks [15, 29]. More precisely, it represents a legal instrument that defines a powerful communication tool, that can be used to determine contractual partners engaged in responsible processes or to incite them to develop this aspect. It is therefore of crucial interest to integrate informational components in the contracting process of the formation of the contract, to motivate and lead to the partners agreement to the concept of social good for the use of AI, as it has already been demonstrated for climate change [31]. In practice, with both public and private institutions, it is feasible to internally integrate the observance of ethical standards in employment contracts. Externally, ones can consider to obtain the adherence of institutional, commercial or financial partners to the defined ethical standards, to ensure the integration of technical standards enabling a responsible development of AI within the partner organizations. The freedom that frames the contractual tools makes it possible to consider adding contractual clauses regulating the use of a research object, or a product, to limit restrictively the utilization and the resale of an AI technology compliantly to social good principles.

4.2 Ensure the use of AI for social good

Despite the fact that the integration of a standardization constitutes an important step toward a responsible use of AI for social good, standards are non-binding and their is no guarantee of their correct application within private or institutional organizations. As a consequence, it is essential to propose tools to ensure the conformity of the acts of an organization that specifically showcases its compliance with a responsible AI program, or to at least inform the public and the stakeholders of the organizations that commit or not. In this extent, new control centers must be developed. In particular, it is important to dissociate the *standardization*, that defines the process of conception and production of reference materials (*e.g* standards) from the *certification* that denotes the conformity assessment obtained by an entity from a third-party body with respect to specific standards [20, 26]. In fact, the latter certification bodies are part of the voluntary standard-setting instruments due to the fact that they choose to comply without any legal obligation. As an example, environmental protection often involves to demonstrate on the packaging of a product its attachment in the fight against climate change, through a label, or by highlighting the efforts made to limit the CO₂ consumption, or even by showcasing its involvement to a sustainable development program aiming at promoting the recourse to local producers [31]. Thereby, it is possible to follow the same approach, but with AI products and toward the use of social good for AI [30].

5 Conclusion

In this paper, it has been shown that the lack of a legal system, despite the crucial need for frameworks and standards regarding AI, leads to the rise of a private standardization. The latter voluntary standardization enables, through the numerous and various tools introduced in this work (summarized in Figure 1 in Appendix), to gradually build foundations for the future legal edifice. Nevertheless, it is crucial to bear in mind that voluntary standards, despite multiple advantages, must remain a temporary alternative or an accompaniment to the adoption of legally binding standards. Indeed, and while they ensure a form of social good for AI, they do not guarantee the respect of the common good, that is protected by governmental standards.

References

- [1] AFNOR, AI standards. <https://norminfo.afnor.org/search?typeResult=normes-etude&commissionID=127690>. Accessed: 2019-08-20.
- [2] Facebook AI ethics institute. <https://www.forbes.com/sites/samshead/2019/01/20/facebook-backs-university-ai-ethics-institute-with-7-5-million/#5e087d8e1508>. Accessed: 2019-08-20.
- [3] G2IA, mandate for the International Panel on Artificial Intelligence. <https://pm.gc.ca/en/news/backgrounders/2018/12/06/mandate-international-panel-artificial-intelligence>. Accessed: 2019-08-20.
- [4] Google AI ethics council. <https://www.theguardian.com/technology/2019/apr/04/google-ai-ethics-council-backlash>. Accessed: 2019-08-20.
- [5] Google to drop Pentagon AI contract. <https://www.washingtonpost.com/news/the-switch/wp/2018/06/01/google-to-drop-pentagon-ai-contract-after-employees-called-it-the-business-of-war/>. Accessed: 2019-08-20.
- [6] ISO, AI standards. <https://www.iso.org/committee/6794475/x/catalogue/p/0/u/1/w/0/d/0>. Accessed: 2019-08-20.
- [7] Montreal declaration for a responsible development of artificial intelligence. <https://www.montrealdeclaration-responsibleai.com>, 2018.
- [8] Bertrand Bellon. Quelques fondements de l'intelligence économique. *Revue d'économie industrielle*, 98(1):55–74, 2002.
- [9] Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In *Advances in Neural Information Processing Systems*, pages 4349–4357, 2016.
- [10] Pavel Castka, Christopher J Bamber, David J Bamber, and John M Sharp. Integrating corporate social responsibility (csr) into iso management systems—in search of a feasible csr management system framework. *The TQM magazine*, 16(3):216–224, 2004.
- [11] Julien Cazala. Le soft law international entre inspiration et aspiration. *Revue interdisciplinaire d'études juridiques*, 66(1):41–84, 2011.
- [12] Olivier Costa, Nicolas Jabko, Christian Lequesne, and Paul Magnette. La diffusion des mécanismes de contrôle dans l'union européenne: vers une nouvelle forme de démocratie? *Revue française de science politique*, 51(6):859–866, 2001.
- [13] Alexander Dahlsrud. How corporate social responsibility is defined: an analysis of 37 definitions. *Corporate social responsibility and environmental management*, 15(1):1–13, 2008.
- [14] Cécilia Darnault. *Les PME face au contentieux économique: essai de guide pratique*. PhD thesis, Aix-Marseille, 2018.
- [15] Gaëlle Deharo. *ingénierie contractuelle et performance de l'entreprise: perspective économique et dynamique de droit des contrats*. 2011.
- [16] Pierre-Marie Dupuy. Soft law and the international law of the environment. *Mich. J. Int'l L.*, 12:420, 1990.
- [17] Nicole Ferry-Maccario, Jan Kleinheisterkamp, François Lenglard, and Nicole Stolowy. *Gestion juridique de l'entreprise*. Pearson Education France, 2006.
- [18] Luciano Floridi, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi, et al. Ai4people—an ethical framework for a good ai society: opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4):689–707, 2018.

- [19] Jean-Gabriel Ganascia. *Le Mythe de la Singularité. Faut-il craindre l'intelligence artificielle?* Le Seuil, 2017.
- [20] Agnès Grenard. Normalisation, certification: quelques éléments de définition. *Revue d'économie industrielle*, 75(1):45–60, 1996.
- [21] Andrew T Guzman and Timothy L Meyer. International soft law. *Journal of Legal Analysis*, 2(1):171–225, 2010.
- [22] Bruce L Hay, Robert N Stavins, and Richard HK Vietor. *Environmental protection and the social responsibility of firms: perspectives from law, economics, and business*. Resources for the Future, 2005.
- [23] Vincent Helfrich. La régulation des pratiques de rse par les normes: le cas de la norme iso 26000 sur la responsabilité sociale. *5e Congrès de l'ADERSE" Transversalité de la RSE: L'entreprise à l'aune de ses responsabilités vis-à-vis de l'homme, de l'environnement et du profit*, 2008.
- [24] Samuel O Idowu, Nicholas Capaldi, Liangrong Zu, and Ananda Das Gupta. *Encyclopedia of corporate social responsibility*, volume 21. Springer New York, 2013.
- [25] Armin Krishnan. *Killer robots: legality and ethicality of autonomous weapons*. Routledge, 2016.
- [26] Andrej Krzan, Sarunya Hemjinda, Stanislav Miertus, Andrea Corti, and Emo Chiellini. Standardization and certification in the area of environmentally degradable plastics. *Polymer degradation and stability*, 91(12):2819–2833, 2006.
- [27] Laure Nurit-Pontier and Stéphane Rousseau. *Risques d'entreprise : quelle stratégie juridique ?* L.G.D.J, 2012.
- [28] Alain Pellet. Les raisons du développement du soft law en droit international: choix ou nécessité?, 2018.
- [29] Eric A Posner. Economic analysis of contract law after three decades: Success or failure. *Yale LJ*, 112:829, 2002.
- [30] David Rolnick, Priya L Donti, Lynn H Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross, Nikola Milojevic-Dupont, Natasha Jaques, Anna Waldman-Brown, et al. Tackling climate change with machine learning. *arXiv preprint arXiv:1906.05433*, 2019.
- [31] Mathilde Sous la direction de Hautereau-Boutonnet and Stéphanie Porchy-Simon. *Le changement climatique, quel rôle pour le droit privé ?* Dalloz, 2019.
- [32] Mariarosaria Taddeo and Luciano Floridi. How ai can be a force for good. *Science*, 361(6404):751–752, 2018.
- [33] Morgane Tual and Vincent Fagot. Intelligence artificielle : ce qu'il faut retenir du rapport de cédric villani. *Le Monde*, 2018.
- [34] UNESCO. Intelligence artificielle: promesses et menaces. *Courrier de l'UNESCO*, (3):70, 2018.
- [35] Rafael Yuste, Sara Goering, Guoqiang Bi, Jose M Carmena, Adrian Carter, Joseph J Fins, Phoebe Friesen, Jack Gallant, Jane E Huggins, Judy Illes, et al. Four ethical priorities for neurotechnologies and ai. *Nature News*, 551(7679):159, 2017.
- [36] Jieyu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez, and Kai-Wei Chang. Men also like shopping: Reducing gender bias amplification using corpus-level constraints. In *Proceedings of the EMNLP*, 2017.
- [37] Jieyu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez, and Kai-Wei Chang. Gender bias in coreference resolution: evaluation and debiasing methods. In *Proceedings of the NAACL*, 2018.

6 Appendix

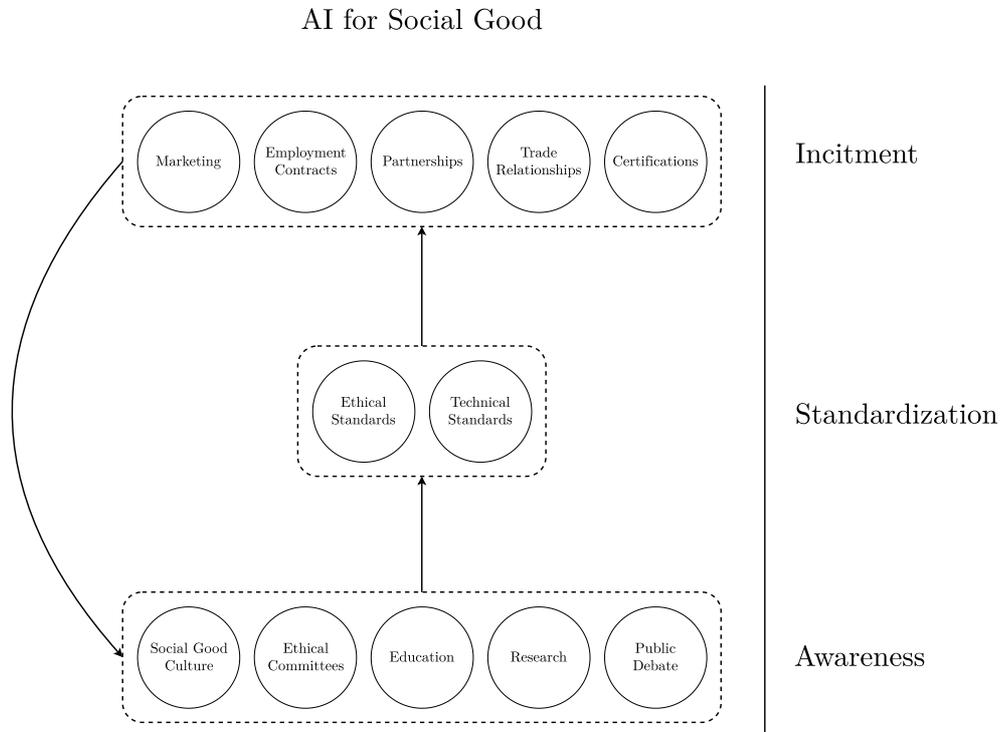


Figure 1: Illustration of the proposed Social Good for AI model. In the same manner as feed-forward recurrent neural networks, our approach relies on different abstraction layers. In fact, each layer defines a high-level concept that bundles multiple tools (*i.e. neurons*) that are connected to the next concept. The outputs obtained are in fact used again to condition the input, describing a virtuous circle.