# Negative Social Impacts of Artificial Intelligence and the Main Mitigation Actions: A Systematic Review

Vanessa Marques Paes(✉), Franciane Freitas Silveira, and Alessandra Cristina Santos Akkari Munhoz

Federal University of ABC, Santo Andre, SP, Brazil
vanessa.paes@ufabc.edu.br

**Abstract.** The challenges faced by those working with Artificial Intelligence (AI) technologies that directly impact human social or ethical life can be divided into two main categories, encompassing philosophical concerns about the feasibility of incorporating ethics into algorithms and the technological challenges of AI development. Among the challenges are the ethical and social impacts connecting two intertwining domains, the human and technology. This study aimed to map the main social impacts of AI and the recommendations to mitigate these impacts, using the systematic literature review (SLR) as a methodological approach. Among the results, it became evident that the research for identification and analysis of social impacts has been growing in recent years, as well as the interest in mitigation actions. Through the analysis of the relationship between these two themes, it was possible to identify that one of the main impacts addressed is related to the fear of technology usurping jobs in the areas of commerce, industry, and transportation, which are the areas that already have practical examples where technology, in fact, is performing human functions. The main mitigation actions for this social impact included increasing public debate about this circumstance and creating laws, principles and methods for regulating the use of AI.

**Keywords:** Artificial Intelligence · Social Impacts · Mitigation

## 1 Introduction

The challenges faced by those working with Artificial Intelligence (AI) technologies that directly impact human social or ethical life can be divided into two main categories: philosophical concerns about the feasibility of incorporating ethics into algorithms, and the technological challenges of AI development [1]. Among the challenges are the ethical and social impacts connecting two intertwining domains: human and technology, because humans are technological beings and technology is a social entity [2]. This means that ethical and social values are not only integrated into systems, but instead are created daily in the interaction between man and machine [2]. Just as knowledge develops into technology and innovation, discussions about the social impacts of these new technologies also need to be assessed prior to their deployment. Such evaluation


L. Pereira et al. (Eds.): IDEAS 2022, DSI, pp. 24–33, 2023.
https://doi.org/10.1007/978-3-031-29129-6_3

deserves to be carried out more intensely, because experts indicate that this technological revolution must be started in the right way, at the risk of the social situation degrading even more [3].

Social impact can be defined as changes in people's way of life in relation to culture, community, political interests, environment, health, well-being, personal rights, fears and aspirations [4]. The concern with the social impacts of the use of AI goes beyond the issues related to deaths resulting from the use of autonomous vehicles, and puts light on the risks involved with the most common use of this technology, such as algorithms that make decisions in selection processes, lawsuits, and even in analysis and release of personal credit, because if there is any error or prejudice inserted into the system, real people will have their lives affected [5].

It is also noteworthy that, since technologies are being implemented faster than the critical analysis of these risks, there are possibly incorrect solutions also being implemented and, consequently, risks that directly affect the people involved.

The literature indicates that the process of rework and mitigation of these impacts could be avoided if the whole procedure of evaluation and critical awareness was implemented from the development project of each system that involves the use of Artificial Intelligence [6].

This reflection has also been driven by the COVID-19 pandemic, which started in China in 2019 and has drastically altered the way people relate to each other and perform day-to-day activities, stimulating the use of various technologies to perform tasks that were previously performed in a face-to-face manner by the majority of the population [7–9].

At a time when discussions about the importance and applicability of ethics have become increasingly important and necessary, especially for AI-oriented applications, it becomes paramount to discuss the social impacts resulting from this new Industrial Revolution, as well as to evaluate the ways to eliminate or mitigate the possible negative impacts generated by such technological advances. These, then, are the objectives of this research, which will undertake a systematic literature review in order to map the negative social impacts of AI and the main mitigation actions.

The definition of AI is complex [10] and the first researcher who sought this definition was Turing. The author defined intelligent behavior as the ability to achieve human-level performance in cognitive tasks, enough to fool an interrogator. The test that the author proposed is that the computer should be interrogated by a human by means of a conversation through a keyboard, and if the interrogator cannot tell whether there is a computer or a human on the other end, the equipment would pass the test [11]. However, this definition does not separate technical knowledge from intellectual perception [10].

One of the main attempts to define AI is the proposal developed by Russell and Norvig and classified into two dimensions, the top and bottom dimensions. According to Table 1, the definitions at the top (systems that think like humans and systems that think rationally) are aimed at the thinking and reasoning processes, while those at the bottom (systems that act like humans and systems that act rationally) are aimed at the behavioral processes of AI systems. The definitions on the left (systems that think like humans and systems that act like humans) measure success in terms of trustworthiness to human performance, while those on the right (systems that think rationally and systems

that act rationally) measure the relation to an ideal compared to a performance measure, called rationality [12] (Table 2).

**Table 1.** AI Definitions

| Systems that think like humans | Systems that think rationally |
|---|---|
| "The exciting new effort to make computers think… Machines with minds, in the full and literal sense." (Haugeland, 1985) | "The study of mental faculties through the use of computer models. ""<br>(Chamiak and 1985) |
| "The automation activities we associate with human thought, activities such as decision making, problem solving, learning…"<br>(Bellman, 1978) | "The study of the computations that enable one to perceive, reason, and act." (Winston, 1992) |
| **Systems that act like humans** | **Systems that act rationally** |
| "The art of creating machines that perform functions that require intelligence when performed by people. ""<br>(Kurzweil, 1990) | "Computational Intelligence is the study of the design of intelligent agents." (Pooleet 1998) |
| "The study of how to make computers do things that, at the moment, people are better at. ""<br>(Rich and Knight, 1991)" | "AI… is concerned with intelligent behavior in artifacts."(Nilsson, 1998) |

Source: Adapted from Russel and Norvig (2002).

**Table 2.** Criteria for article selection

| | Scopus | Web of Science |
|---|---|---|
| Formula for advanced search | TITLE-ABS-KEY = ("ARTIFICIAL INTELLIGENCE" OR" AI) AND (ETHIC * OR MORAL *) AND (SOCI *) | TS = (("ARTIFICIAL INTELLIGENCE" OR AI) AND (ETHIC * OR MORAL *) AND SOCI *) |
| | TITLE-ABS-KEY = standard search fields (title, abstract, and keywords) | TS = standard search fields (title, abstract, and keywords) |
| Comments | AND = Boolean operator | AND = Boolean operator |
| | OR = Boolean operator | OR = Boolean operator |
| | SOCI = the word were used incomplete to cover social and society words | SOCI = the word were used incomplete to cover social and society words |

Source: Own authorship (2021).

There is a tension between the human-centered approach and the rationality-centered approach. The human-centered approach must be an empirical science, involving hypothesis and experimental verification. The rationalist approach involves a combination of mathematics and engineering [13].

## 2 Methodological Approach

A systematic literature review (SLR) was used to seek answers to the following questions:

– What are the main social impacts tied to the use of AI?
– What are the main recommendations to eliminate or mitigate these impacts?

The RSL was developed in three main stages, the first being the planning stage, in which the need for the research was identified. The second stage of the RSL involved conducting the review, which was accomplished by searching the publications, selecting the articles according to the established criteria, extracting the data, validating the quality of the publications, and synthesizing the data. The final step involves dissemination of the data [14].

The academic research bases used were the Web of Science and Scopus platforms. The search filters were selected by keeping only scientific articles, of which the publications were made between the years 2000 and 2020, that the language of the publication was in English and from the areas of computer science, social sciences, engineering, business arts and humanities.

Based on the 690 publications selected from the main scientific research databases, as described in Table 1, a systematic literature review was developed. After obtaining the 690 publications, the eligibility criteria were applied, divided into three filters.

The first filter was the analysis of the titles. In this step, all titles and keywords used per author were evaluated, in order to validate the content of the article, which should be related to research and identification of principles and ethical impacts of the use of AI. The second filter was the analysis of the abstracts and conclusions. In this process, the abstracts of the publications were analyzed in order to further evaluate the objective and results of the article. The last filter used was the complete analysis of the publication. In this step, 175 articles were fully read in order to qualitatively identify the main social impacts of AI use, as shown in Fig. 1. All the mentioned steps were performed manually using Microsoft Excel.

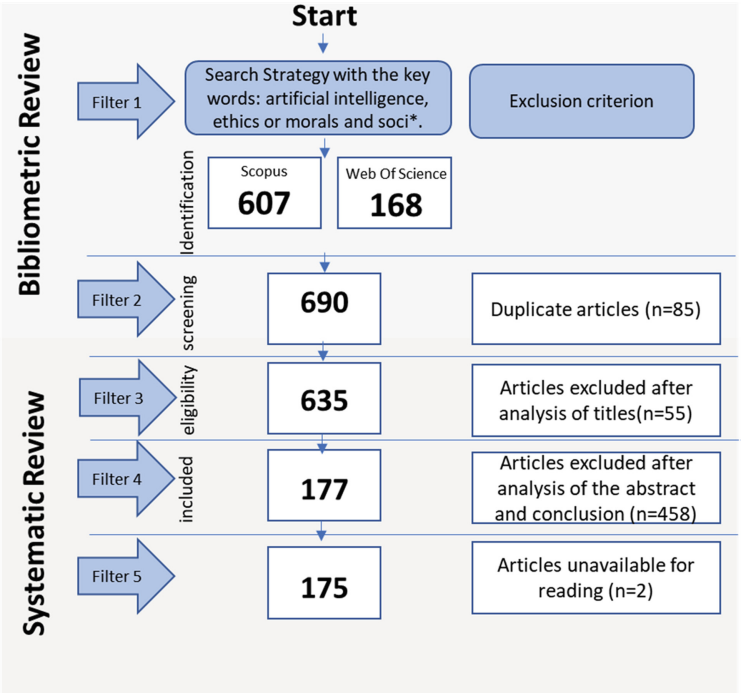


**Fig. 1.** Flowchart of Articles Eligibility Source: Own authorship (2021)

The distribution of publications over the years shows that this is a topic on the rise, as shown in Fig. 2.

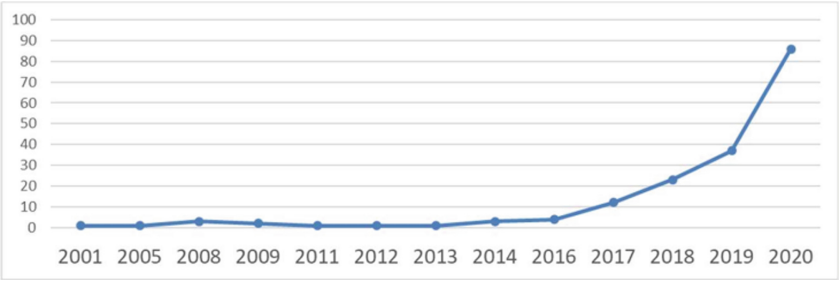


**Fig. 2.** Publication per year Source: Own authorship (2021)

One of the limitations of this study is related to the focus of AI impacts on society as a whole, this means that impacts focusing on individuals were not mapped or developed during the proposed analyses.

## 3 Results and Discussion

Regarding RSL analysis procedures, the articles evaluated were categorized according to affinity and according to the areas of AI use, so that twelve categories of analysis were identified, as follows:

i) government and policy
ii) Society
iii) Leisure
iv) Informatics and technology
v) Industry
vi) Agriculture
vii) Services
viii) Transport and logistics
ix) Commerce
x) Military
xi) Health
xii) Education

After the first categorization, the results were again distributed in nine categories of social impact, as follows:

i) Reduction of autonomy and responsibility: the decision to perform certain activities no longer requires human interference
ii) Usurpation of jobs by automation: formal jobs performed completely by machines or systems
iii) Over-dependence on technology: increased levels of stress due to over-use.
iv) Lack of transparency: uncertainty about decision criteria.
v) Environmental impacts: damage to nature caused by technology
vi) Injustice: unfair results
vii) Bias and discrimination: results based on biases of the historical data or the creators of the systems
viii) Data breach: sensitive data made available
ix) Risk of injury: harm to the human body or life

In this study, social impacts were classified into nine categories of analysis, according to the citations of the publications analyzed, and the amount of citations of each impact will be presented title by title, showing how many times this theme was addressed among the 175 articles evaluated.

In order to identify how the social impacts are distributed among the main areas of AI use, a matrix was prepared to identify the relation between social impacts and the areas where AI is used (Fig. 2). To obtain these data, all articles in this study were manually tabulated according to the categories of the systematic literature review, and each time an article mentioned an impact and an area of action, these were identified in a table.

After analyzing the publications, it was found that, despite being a relevant subject today, environmental impacts have the lowest percentage of citation, reaching at most 2%

in the areas of government and society. The impact related to prejudice and discrimination is present in all categories, with citation percentages higher than 19%, according to Fig. 3.

In Fig. 3, it was noted that there is a strong relationship between the impact related to job usurpation and the areas of industry and services. This is due to the fact that such organizations are rapidly deploying AI to manage their processes and employees [15]. Thus, the overuse of automation and systems capable of performing complex tasks can create a gap in the operational workforce [16] and the use of artificial intelligence is directly linked to the increase in unemployment rates, especially in activities that require a low level of education [17]. The social implications of AI are complex, and especially in relation to jobs related to the automation of manual or digital tasks, there is a risk that jobs will be directed to a group with greater technical ability, and for this reason it is necessary to create labor policies that can ensure the sustainability of all levels of jobs. [18].

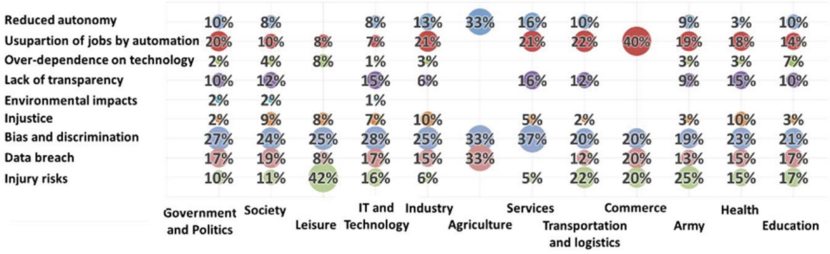


**Fig. 3.** Relationship between Social Impacts and AI Usage Area Source: Authors (2021)

After this step, the data, the results were again distributed in eleven categories of mitigation actions, as follows:

i) Standardization of the AI language, translating ethical principles into protocols
ii) Laws, principles, regulations or policies to ensure ethical implementation IA
iii) Breadth and depth of the public debate
iv) Translation of the main studies and reports.
v) Alternative continents for major conferences on AI research and on ethics and governance
vi) Increasing the accessibility of correct information to the public in the form of fact sheets and statements of ethical values on reliable web pages
vii) Establish joint programs and/or trainings
viii) Fees for the use of AI
ix) Creation of legally deregulated, or special zones, for research and development in robotics
x) Researchers and programmers should seek advice from ethicists to avoid mistakes (multidisciplinary)
xi) Preserving human decision-making processes for certain types of decisions

In the last stage of the analysis, the main recommendations for the elimination or mitigation of impacts found in the selected literature were analyzed, with the aim

of identifying which actions have greater relevance in the publications evaluated, and identify the relationship with the social impacts mentioned. In this sense, it is identified that the most comprehensive action is related to the need to "increase the breadth and depth of public debate about the risks that the AI" can provide. Such action was evident in all categories of social impacts with percentages higher than 20%. Encouraging the understanding and sharing of information regarding AI is an urgent challenge for global society [19], and seeking to align information is a necessary and challenging task [20],,as shown in Fig. 4.

Secondly, the action of "creating laws, principles and regulation of AI use" was identified, which also received mention in all social impact categories. Moreover, in third place, the indication of "standardize AI language" with percentage results above 9%. The first three recommendations added up to 66% of the total. One of the caveats with the lowest adherence is the creation of "fees or taxes for companies that use AI systems to replace humans," with a percentage of 0.3%. The action of "increasing the breadth and depth of discussions regarding AI" highlights the importance of public knowledge, and it should be noted that this debate can bring out both the most optimistic [21] and the most pessimistic opinions [22].

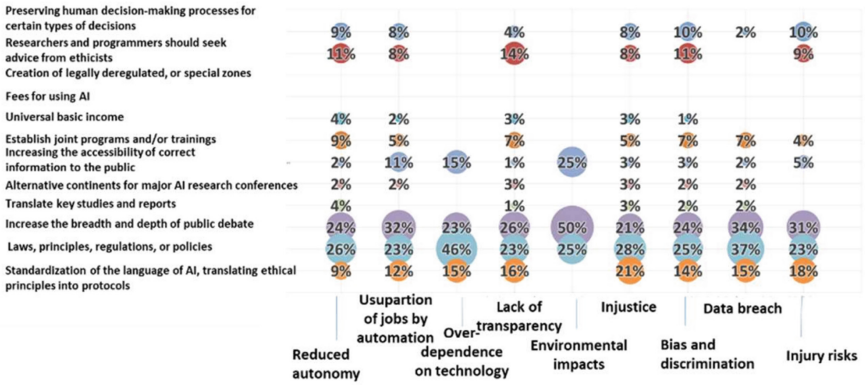


**Fig. 4.** Impacts x mitigation actions relationship Source: Authors (2021)

It is worth noting that despite the negative impacts, AI brings numerous benefits to society, a study conducted in the United States reported that over a 30-year period in The New York Times newspaper, the amount of articles that addressed AI in a positive way was consistently more positive than negative[23]. However, in recent years evaluated, it was indicated that there was an increase in reports with higher levels of concern [23].

The public debate about the social impacts of AI is in its early stages [21], and needs to be developed in a balanced way, taking into account opportunities and risks, in order to enable initiatives that foster innovation but also exercise control over potential risks [24].

## 4 Conclusion

This article showed that the research aimed at identifying social impacts has been growing in recent years, as well as the interest in mitigation actions. Through the analysis of the relationship between these two themes, it was possible to identify that one of the main impacts addressed is related to the fear of technology usurping jobs in the areas of commerce, industry, and transportation, which are the areas that already have practical examples where technology is actually performing human functions, and the main mitigation actions for this social impact is the "increase of public debate about this circumstance" and the "creation of laws, principles and methods of regulation of the use of AI", because in this way the understanding about the benefits versus the impacts will be understood and through this process, laws and regulations can be defined with the goal of mitigating them. Therefore, of the eleven mitigation actions addressed in the articles evaluated in this study, these are the two main actions, and through them the nine impacts evaluated can be mitigated. In this direction, it is evident the need for further research and relationship analysis of these two mitigation actions, in order to identify how to apply them effectively.

**Acknowledgments.** The authors would like to thank Federal University of ABC and the Engineering and Innovation Management graduate program for the opportunity and support. This research was supported by the EP32280009/3006.0082.03/202–41.

## References

1. Vanclay, F.: International principles for social impact assessment. Impact Assess Proj. Apprais **21**, 5–12 (2003)
2. Floridi, L., et al.: AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Mind. Mach. **28**(4), 689–707 (2018). https://doi.org/10.1007/s11023-018-9482-5
3. Anderson, M., Anderson, S., Armen, C.: Towards machine ethics: implementing two action-based ethical theories. In: Proceedings of AAAI 2005 Fall Symposium Machine Ethics, pp. 1–7 (2005)
4. Verbeek, P-P.: Designing the morality of things: the ethics of behaviour-guiding technology. Des. Ethics, 78–94 (2017)
5. Vesnic-Alujevic, L., Nascimento, S., Pólvora, A.: Societal and ethical impacts of artificial intelligence: critical notes on European policy frameworks. Telecomm. Policy **44**, 101961 (2020)
6. Fachin, P.: Novo Ciclo Tecnológico Requer que a Sociedade Repense seu Pacto Fundador. http://www.ihu.unisinos.br/159-noticias/entrevistas/595904-novo-ciclo-tecnologico-requer-que-a-sociedade-repense-seu-pacto-fundador-entrevista-especial-com-glauco-arbix. Accessed 21 Nov 2019
7. Cheng, S.O., Khan, S.: Europe's response to COVID-19 in March and April 2020 - a letter to the editor on World Health Organization declares global emergency: a review of the 2019 novel coronavirus (COVID-19) (Int J Surg 2020;76:71–6). Int. J. Surg. **78**, 3–4 (2020)
8. Hiscott, J., Alexandridi, M., Muscolini, M., et al.: The global impact of the coronavirus pandemic. Cytokine Growth Factor Rev. **53**, 1–9 (2020)

9. Pan, X-B.: Application of personal-oriented digital technology in preventing transmission of COVID-19. Ir. J. Med. Sci. **189**, 1145–1146 (2020)
10. Dobrev, D.: A Definition of Artificial Intelligence (2012)
11. Turing, A.M.: On computable numbers, with an application to the Entscheidungsproblem. Proc. London Math. Soc. s2–42, 230–265 (1937)
12. Norvig, P., Russel, S.: Artificial Intelligence. A Modern Approach, Prentice Hall Press., Upper Saddle River, NJ, USA (2020)
13. Kahneman, D., Tversky, A.: The psychology of preferences. Sci. Am. **246**, 160–173 (1982)
14. Wahono, R.S.: A systematic literature review of software defect prediction. J. Softw. Eng. **1**(1), 1–16 (2015)
15. Hughes, C., Robert, L., Frady, K., et al.: Artificial intelligence, employee engagement, fairness, and job outcomes. In: Managing Technology and Middle- and Low-skilled Employees, pp. 61–68. Emerald Publishing Limited, (2019)
16. No TALOTAIBI SS: Ethical issues and related considerations involved with artificial intelligence and autonomous systems. Int. J Adv. Comput. Sci. Appl. **9**, 35–40 (2018)
17. Bordot, F.: Artificial intelligence, robots and unemployment: evidence from OECD countries. J. Innov. Econ. Manag. **37**(1), 117–138 (2022)
18. Tschang, F.T., Almirall, E.: Artificial intelligence as augmenting automation: implications for employment. Acad. Manag. Perspect. **35**(4), 642–659 (2021)
19. ÓhÉigeartaigh, S.S., Whittlestone, J., Liu, Y., Zeng, Y., Liu, Z.: Overcoming barriers to cross-cultural cooperation in AI ethics and governance. Philos. Technol. **33**(4), 571–593 (2020). https://doi.org/10.1007/s13347-020-00402-x
20. Vrščaj, D., Nyholm, S., Verbong, G.P.J.: Is tomorrow's car appealing today? ethical issues and user attitudes beyond automation. AI Soc. **35**(4), 1033–1046 (2020). https://doi.org/10.1007/s00146-020-00941-z
21. Ouchchy, L., Coin, A., Dubljević, V.: AI in the headlines: the portrayal of the ethical issues of artificial intelligence in the media. AI Soc. **35**(4), 927–936 (2020). https://doi.org/10.1007/s00146-020-00965-5
22. Shariff, A., Bonnefon, J.-F., Rahwan, I.: Psychological roadblocks to the adoption of self-driving vehicles. Nat. Hum. Behav. **1**, 694–696 (2017)
23. Fast, E., Horvitz, E.: Long-term trends in the public perception of artificial intelligence. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 31, no. 1 (2017)
24. Chuan, C-H., Tsai, W-HS., Cho, S.Y.: Framing artificial intelligence in American Newspapers. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, pp. 339–344 ACM, New York, NY, USA (2019)